The Role of Mathematics in Physical Sciences

Interdisciplinary and Philosophical Aspects

Edited by

Giovanni Boniolo, Paolo Budinich and Majda Trobok

Springer

THE ROLE OF MATHEMATICS IN PHYSICAL SCIENCES

The Role of Mathematics in Physical Sciences

Interdisciplinary and Philosophical Aspects

Edited by

GIOVANNI BONIOLO
University of Padova, Italy

PAOLO BUDINICH
The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy

and

MAJDA TROBOK
University of Rijeka, Croatia

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 1-4020-3106-8 (HB)
ISBN 1-4020-3107-6 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Printed on acid-free paper

Printed in the Netherlands.

Contents

Contributing Authors

Vladimir I. Arnold
Steklov Mathematical Institute
8, Gubkina Street
119991 Moscow - Russia
arnold@mi.ras.ru

Giovanni Boniolo
Dipartimento di Filosofia
Università di Padova
Piazza Capitaniato 3
35100 Padova – Italy
giovanni.boniolo@unipd.it

Paolo Budinich
The Abdus Salam International Centre for Theoretical Physics
Strada Costiera 11
34014 Trieste - Italy
fit@ictp.trieste.it

Lucio Crivellari
INAF – Osservatorio Astronomico di Trieste
Via G.B. Tiepolo 11
I-34131 Trieste – Italy
Instituto de Astrofísica de Canarias
c/. Via Láctea S/N
E-38200 La Laguna, Tenerife - Spain
crivellari@ts.astro.it

Mauro Dorato
Dipartimento di Filosofia
Università di Roma 3
Via Ostiense 234
00146 Rome – Italy
dorato@uniroma3.it

Dennis Dieks
Institute for the History and Foundations of Science
Utrecht University
P.O. Box 80000
NI 3508 TA Utrecht – The Netherlands
dieks@phys.uu.nl

Giancarlo Ghirardi
The Abdus Salam International Centre for Theoretical Physics
Strada Costiera 11
34014 Trieste – Italy
INFN – sez. Trieste – Italy
ghirardi@ictp.trieste.it

Giulio Giorello
Dipartimento di Filosofia
Università di Milano
Via Festa del Perdono
20100 Milano – Italy
g.giorello@tiscalinet.it

Victor Gómez Pin
Universitat Autònoma de Barcelona
Barcelona – Spain
operatal@terra.es

Miklós Rédei
Department of History and Philosophy of Science
Faculty of Sciences
Loránd Eötvös University
P.O. Box 32
H-1518 Budapest 112 – Hungary
redei@hps.elte.hu

Andrés Rivadulla
Universidad Complutense de Madrid
Dpto. de Lógica y Filosofia de la Ciencia
28040 Madrid – Spain
arivadulla@filos.ucm.es

Navjyoti Singh
Indian Institute of Advanced Studies
Rashtrapati Nivas
Shimla 171 005 – India
navjyoti@rediffmail.com

Corrado Sinigaglia
Dipartimento di Filosofia
Università di Milano
Via Festa del Perdono
20100 Milano – Italy
corrado.sinigaglia@unimi.it

Michael Stöltzner
Institute for Science and Technology Studies
University of Bielefeld
P.O. Box 100131
D-33501 Bielefeld – Germany
stoeltzner@iwt.uni-bielefeld.de

Zvonimir Šikić
Department of Mathematics
University of Zagreb
Bijenicka 30
10000 Zagreb – Croatia
zsikic@math.hr

Nikola Zovko
Rudjer Boskovic Institute
Bijenicka 54
10000 Zagreb – Croatia
zovko@thphys.irb.hr

Preface

What the role of mathematics in physical sciences is, is a relevant philosophical and historical question whose answer is necessary to fully understand the real status of physics, in particular of contemporary physics.

Exactly the wish to have good and plausible answers has spurred physicists, mathematicians, historians of science, and philosophers of science from many countries to join together and friendly but rigorously discuss. From that meeting, which was held in the wonderful Isle of Losinj (Croatia) in 2003, this book had its origin.

Actually, it does not simply contain the text of the lectures given. It is something different and something more. Some chapters are new and improved versions of what was presented. Some others have been added to enrich the variety of possible suggestions.

This book has been published in occasion of the 40th anniversary celebrations of the Consorzio per la Fisica of Trieste.

The editors

Acknowledgments

We would like to thank those who have made possible the meeting in Losinj and consequently this book, by sponsoring and financing them. They are: the Italian Ministery for Foreign Affair; the Consorzio per la Fisica, in the person of their president, prof. Franco Bradamante, and subsequently prof. Giancarlo Ghirardi; the Area Science Park of Trieste; the International School for Advanced Studies; the "Abdus Salam" International Centre for Theoretical Physics; the Osservatorio Astronomico of Trieste; the Università Popolare of Trieste; the Istituto Italiano di Cultura of Zagabria and the Ruder Boskovic Institute of Zagabria. And last but not least, the wonderful staff, headed by dr. Aldo Baldini, which, with unlimited, enthusiastic dedication, rendered fluent and enjoyable the meeting ands less tiring the work on the book.

The editors

THE ROLE OF MATHEMATICS IN PHYSICAL SCIENCES – INTERDISCIPLINARY AND PHILOSOPHICAL ASPECTS

Introduction

GIOVANNI BONIOLO[1], PAOLO BUDINICH[2], MAJDA TROBOK[3]
[1]*University of Padova, Padova, Italy;* [2]*The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy;* [3]*University of Rijeka, Rijeka, Croatia*

As only a cursory examination of the subject can illustrate, mathematics and physics have been related for centuries and now it seems quite impossible to think the latter without the former. In other words, to speak about the indispensability of mathematics for physics appears to be a real platitude. However it is not at all that simple and unproblematic. In fact a lot of problems arise from this relation: is mathematics really indispensable for physics, or could we have physics without mathematics? Did physics without mathematics exist? Could physics without mathematics exist now? Which are the relations between physics and mathematics? Is mathematics just a tool, or something more? Is it the language in which is written the nature or is it the language by means of which we try to know nature? Has it only a role in the logical structuration of a physical theory or does it furnish also a good path to discover new physical entities? Should we think physically and then should we add the mathematics apt to formalise our physical intuition, or should we think mathematically and then should we interpret physically what found? Can physics generate new mathematics? Can mathematics generate new physics? How can we explain the success of mathematics in the physical sciences? Should it really be explained, or is such a question a pseudo-question? Are there any limits to the mathematical applications? Does a pure mathematical method to construct new physical theories exist? Do we get mathematical objects by abstraction from real objects, or are they a direct product of our intuition?

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 5–8.

All these questions and problems have been discussed in this book from different perspectives and by authors with different philosophical backgrounds.

We have thought of dividing the book into three parts. The first one contains four contributions on the historical role of mathematics in physics.

Giorello and **Sinigaglia** question the idea that mathematical objects are not obtained by abstraction from real ones, but rather that they are generated by mathematical practice. The authors analyse this thesis in the light of two historical cases: the evolution of complex numbers and the development of Heaviside’s Operational Calculus and give arguments for supporting Lakatos’s idea of quasi-empiricism in mathematics.

Gómez Pin discusses the problem of the ontological priority between continuous and discrete quantity and analyses the relationship between discrete and continuous quantity as one of the main topics in both history of philosophy and science. He explains that, while the unit of discrete quantity is a genuine (atomic) unit but ontologically is a vacuum, the unit of a continuous quantity has great ontological weight but it is in fact a false (non atomic) unit. The history the author concentrates on is the debate Aristotle-Thom/Dedekind-Cantor.

Rédei presents J. von Neumann’s view on mathematical and axiomatic physics. The author argues that the common evaluation of von Neumann’s view on the mathematical rigour in physics, according to which he considered the axioms of set theory as a purely formal system, is misleading. Namely, as the author points out, von Neumann thought that conceptual clarity and an intuitively satisfactory interpretation was more important for a physical theory than its mathematical rigour and precision.

Finally, **Singh** looks at the Indian tradition of mathematics with respect to theories of mind and matter. In particular, the author explores the reason for the absence of mathematical physics in Indian mathematical traditions, while at the same time the mathematical thought was employed by several Indian philosophical schools in order to understand the functioning of human’s mind. The author enquires the reasons for this analysing the connection between mathematics and the idea of causation in Indian tradition. The relation between causation and mathematics is clarified through the causal analysis of numeric cognition.

The second group of papers deals with philosophical analyses on the interaction between mathematics and physics.

Boniolo and **Budinich** join the contemporary discussion about the relation between mathematics and physics, *via* a semiotic approach, which is useful for the many aspects it allows us to tackle. In particular, they argue that the problem of the effectiveness of mathematics in physics is actually a false problem, caused by a misunderstanding of contemporary theoretical

physics, which is intrinsically mathematical. Finally, they emphasize what they call Dirac's methodological revolution according to which the contemporary physical theory should be constructed by working with pure mathematics instead of reflecting conjecturally only on physical phenomena, thus allowing the discovery of new phenomena, as it happened with the discovery of antimatter, gravitational lenses and so on.

Crivellari looks at the algorithmic representation of astrophysical structures and presents an iterative structural algorithm that is the numerical stimulation of the physical processes that occur in a stellar atmosphere. Through its analysis the author tries to show that, when the right mathematics is to be determined, it is the physics of the problem to have a bearing on what the most efficient solution is.

Dieks discusses the, so called, unreasonable effectiveness of mathematics and argues that, quite the contrary, its effectiveness is actually to be expected and its being unreasonable is unfairly attributed to it. Dieks shows that mathematics is flexible and versatile and that it is the very difference in nature between mathematics and physics that makes it applicable in the most disparate scientific domains and hence vastly effective. The author illustrates his view by offering many examples from fundamental physics.

Dorato questions the mathematical aspects of physics, by analysing the possible connection between the problem of effectiveness of mathematics in the natural sciences and the philosophical questions concerning the nature of natural laws. The author argues that the problem of the effectiveness is, contrary to what some authors endorse, a genuine one and criticises the algorithmic conception of law. The aim is to review and evaluate the available literature on that matter and suggest new possible directions of inquiry regarding the problem.

Ghirardi analyses some mathematical aspects of modern science and argues that new and inexplicable phenomena can suggest new and innovative theoretical and mathematical perspectives; those perspectives and their formal aspects might in turn yield new and innovative views about nature, and therefore all such formal aspects should be fully developed whenever they qualify themselves as successful tools, to account for some basic features of a revolutionary phenomenological framework.

Rivadulla presents some theoretical explanations in mathematical physics in the context of the analysis of the problem of the usefulness of mathematics in physics. The authors criticises the view according to which mathematics tallies with nature since it is a structural science as nature is, and because of some evolutionary reasons that make us adapted to the structured world; Rivadulla gives reasons for sustaining that such a view is incomplete because it does not take into account the overdetermination of physics.

Šikić is interested in the relationship among mathematics, physics and music. He investigates the Pythagorean law of small numbers and its relevance in order to interpret our sensory discriminations of consonance *vs*. dissonance. The author argues that the view, which is allegedly confirmed by the fact of non-western musical traditions, according to which we should take the discriminations to be acquired and subjective, is a wrong one.

Finally, **Stöltzner** looks at theoretical mathematics and points to the philosophical significance of the Jaffe-Quinn debate, which is viewed as a paradigm for problems of rigour and mathematical ontology. After going over the essential of the debate, the author concentrates on the quasi-empirical character of mathematics and the dialectics of proofs and refutations, trying to make sense of "theoretical mathematics" within the Lakatosian approach.

The third part of the book contains two interesting considerations on the relation between mathematics and physics that spur us to think about it in a wider way.

In particular **Arnold** joins the discussion about the relationship between mathematics and physics. He presents, through examples, the problem of the mathematical rigour of the bases of physics and explains what the utility of a precise mathematical perspective of the real world is. The author also offers some arguments for the existing difference in the approach to the truth as understood by mathematicians and physicists.

In the last paper, **Zovko** questions the notions of value and meaning in quantum universe. The author suggests that the mental universe is subject to the same mechanism as the physical universe and that human thoughts are just actual quantum events over the entire brain or over a large part of it. He points out that both the mental and material universe can be unified as a physical reality on a deeper level, beyond our direct experience; such a realm could also accommodate ethical concepts of choice, meaning and value.

PART 1

MATHEMATICS AND PHYSICS: REFLECTING ON THE HISTORICAL ROLE OF MATHEMATICS

OLIVER HEAVISIDE'S "DINNER"

*Algebraic Imagination and Geometrical Rigour**

GIULIO GIORELLO[1] and CORRADO SINIGAGLIA[2]
[1]University of Milan, Milan, Italy; [2]University of Milan, Milan, Italy

Abstract: In the following pages we begin, in the first chapter, with a reappraisal of some ideas of Edouard Le Roy about mathematical experience, mainly in relation with the history of complex numbers. In the second chapter we discuss in some detail the *i*-story, and we draw a comparison between "Imaginary Quantity" and Operational Calculus from the perspective of Heaviside's conceptions of the growth of mathematics. In the third chapter we reconstruct the δ-story, i.e. the Heaviside calculus leading to the *constitution* of a new mathematical object, the so-called Dirac's δ-function. Finally, in the last chapter, we bring together methodological and historical considerations in order to support Lakatos' idea of *quasi-empiricism* in mathematics.

Key words: complex numbers; operational calculus; δ-function; abstraction; quasi-empiricism in mathematics; mixed mathematics; applications to physics.

1. "MATHEMATICAL FACTS" AS CONSTRAINTS

> Le progrès [de la Mathématique] consiste moins en une application de formes intelligibles données d'avance rigides et toutes faites qu'en une création incessante de formes intelligibles nouvelles, en un élargissement graduel des conditions de l'intelligibilité. Elle suppose une transformation de l'esprit lui-même. (Le Roy, 1960, p. 304).

* We wish to thank G. Bertolotti, G. Boniolo, P. Budinich, V. Fano, N. Guicciardini, and B. Sassoli for suggestions and comments.

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 11–32.

The quotation is from Le Roy's lectures at the Collège de France (Paris) in the years 1914-1915 and 1918-1919. More or less in the same years, Le Roy's key idea is echoed in Pierre Boutroux's search for the objective character of mathematical knowledge, based on

1. the so called "résistance" (resistance) of the mathematical matters to our will (we have really some "mathematical facts") and
2. the "contingence" (contingency) of mathematical findings or discoveries (see e.g. (Boutroux, 1920)).

Le Roy's version, as we shall see, helps to clarify crucial epistemological notions concerning "discovery/invention" in mathematics, mainly in connection with Lakatos' quasi-empiricism (Lakatos, 1976a); see also (Crowe, 1975; Gillies, 2000; Cellucci, 2000). Moreover, even if the title of Le Roy's lectures sounds *Pure Mathematical Thought*, some of his remarks contribute powerful insights into the standard dichotomy pure/applied mathematics, and throw important light on the controversial matter of the status of "mathematical objects". Indeed, in Le Roy's own words (Le Roy, 1960): "Même en Analyse pure, l'expérience joue un rôle, et un rôle capital. L'invention y est souvent découverte" (p. 298); see also (Hadamard, 1949).

According to Le Roy (see Boutroux point (1)), the working mathematician receives some inputs from the constellation of established ideas; however this constellation is not sufficient for generating outputs. The case of complex numbers will be exemplar. Le Roy observes (Le Roy, 1960):

> Les [quantités] imaginaires ne se *déduisent* pas de la science antérieure. Mais elles sont *réclamées* par celle-ci comme une condition de sa vie et de son progrès (p. 298).

He goes on:

> [Les quantités imaginaires] marquent pour l'analyste je ne sais quelle *obligation* de synthèse créatrice. Et leur apparition au bout d'une foule de voies dialectiques diverses, comme point de concours ou centre de convergence, comme élément simple ou invariant méthodique, leur confère une réelle *objectivité*, c'est-à-dire une existence indépendante de nos procédés d'étude. Mais une véritable *expérience* en a été nécessaire pour en arriver là. […] On […] saisira mieux encore [ça] en songeant aux deux problèmes que soulève encore de nos jours – au moins en quelque mesure – la conception des imaginaires. Comment, inventées qu'elles furent pour la résolution de l'équation du second ou du troisième degré, sont-elles non seulement nécessaires, mais encore suffisantes, pour la démonstration générale du théorème de D'Alembert qui domine toute l'algèbre? Comment ne faut-il pas des imaginaires nouvelles pour chaque

degré nouveau d'équation? Pourquoi d'autre part, couples numériques représentables par des vecteurs dans un plan, ne se prêtent-elles à aucune extension, complexes à *n* éléments, vecteurs de l'espace à trois dimensions ou même de l'hyperespace, qui respecte la permanence des formes opératoires? (p. 298).

In these two passages, Le Roy emphasizes the *need* (this is the meaning of the French "*réclamées*") of resorting to a sort of experience in connection with the genesis of objectivity: in his own example, such is the research on factorisation of extensions of **Q** or **R** via some particular complex numbers (e.g. see (Ellison, 1978), as well the research on extensions of **C** violating some relevant formal properties (as in the case of William Rowan Hamilton's quaternions; see (Kline, 1972; van der Waerden, 1985).

So far, so good. However, it is not so easy to find any "counterpart in nature" for complex numbers (Giusti, 1999). This is not tantamount to claiming that complex numbers have no applications to the physical world. Of course, they do; indeed, applications in Electromagnetism and in Quantum Mechanics are well known. The point is rather this: the *genesis* of complex numbers theory, and in the building of the complex functions theory, "abstraction from physical objects" does not seem to be working (Giusti, 1999).

Yet, even here, we are dealing with what Le Roy calls "experience" (Le Roy, 1960):

les imaginaires ne sont pas [...] le résultat d'une création factice. Elles ont été suggérées, amenées, appelées par toutes sortes d'exigences préalables. De bien des manières, avant même qu'on en eût élucidé la théorie, elles voulaient être, elles s'imposaient. Puis elles se sont montrées infiniment fécondes et, de plus en plus à mesure qu'on les expérimentait davantage, elles ont heureusement réagi sur le système entier de la mathématique. Aurait-on pu prévoir *a priori* qu'elles permettraient de résoudre les équations de tous les degrés, qu'elles engendreraient la théorie générale des fonctions par où l'Analyse a été plus que doublée ? Qui aurait pu deviner avant toute expérience le line merveilleux qui devait s'établir entre les nombres e et π et l'unité imaginaire i ? Remarque sur l'imprévisibilité du fait que les imaginaires seraient suffisants pour les équations de tous les degrés, alors qu'on avait démontré l'impossibilité d'une résolution algébrique. De même, qui aurait pu deviner avant toute expérience tant de liens merveilleux entre des éléments réels, établis par l'intermédiaire des nombres complexes? Remarque sur l'étonnement qu'on éprouve à trouver la dépendance foncière de certaines intégrations par rapport aux fonctions de variable imaginaire, jusqu'en physique mathématique. Cauchy a eu profondément

> ce sens du réel dont je parlais tout à l'heure, et le travail par lequel s'est constituée peu à peu la doctrine des imaginaires nous présente vraiment l'aspect d'une élaboration expérimentale. (pp. 301-302)

Let us take an example. Remember that in the ring of the whole numbers **Z** we have the fundamental theorem of arithmetic (a generalization of *Euclid's Elements*, IX, 14: "If a number be the least that is measured by prime numbers, it will not be measured by any other prime number except those originating measuring it" (Euclid, 1956); see also (Heath, 1981)) stating that (except for + 1 and – 1) a number can only be resolved into prime factors in one way. After Pierre de Fermat and mainly thanks to Leonhard Euler, it was an interesting new mathematical practice to study "numbers" of the form $a+\sqrt{D}$, with $a,\ b\ \in$ **Z**, where D is a given integer (positive or negative) which is not a perfect square. The idea was to build a kind of arithmetic of *numeri surdi*; indeed, for $D < 0$, "numbers" $a + b\sqrt{D}$ are complex numbers, as it happens in Euler's procedure for Fermat's equation $x^3 + y^3 = z^3$, where $D = -3$. Moreover, rings **Z**[$\sqrt{D}$] proved to be very useful tools in dealing with many mathematical problems in 19th Century; the same is true for rings **Z**[ζ], where ζ is a complex nth-root of the unity (i.e. $\zeta^n = 1$). Yet, the initial approach to problems like higher forms of Fermat's Last Theorem was guided by the idea that, for **Z**[$\sqrt{D}$] or **Z**[ζ],we have "natural" analogues of Euclid fundamental theorem of arithmetic. Now, this is obviously true for **Z**[$\sqrt{-3}$], but it is false in general. For instance, assume $D = -5$, and try with "numbers" $a + b\sqrt{-5}$, with $a,\ b \in$ **Z**. Check that $6 = 2 \times 3 = (1 + \sqrt{-5})\ (1- \sqrt{-5})$. It proves that in **Z**[$\sqrt{-5}$] it is impossible to get a unique prime factors decomposition. Likewise, it is possible to find counterexamples to the unique decomposition also in **Z**[ζ]. (For the question see (Ellison, 1978, pp. 172-193); see also (Ribenboim, 1979; Giorello and Sinigaglia, 2001)

The *proof* that for some rings unique decomposition does not hold amounts to a *refutation* of this initial conjecture, which seemed so useful within Euler's approach. It is precisely a conjecture like this that for Le Roy (Le Roy, 1960) constitutes a kind of guiding ideas, a sort of preconceived hypotheses, something similar in the realm of mathematics to the empirical hypotheses "qui, selon Claude Bernard, constituent le premier moment d'une expérience" (p. 299). As it is the case of **Z**[$\sqrt{-5}$], we ignore *a priori* wheter or not this conjecture might be incorporated into the body of formal mathematics. The only way to settle the question is (Le Roy, 1960):

> mettre en pratique, en service, mettre à l'essai, faire fonctionner le concept et voir comment il se comporte dans le calcul, bref éprouver l'idée par ses fruits (p. 299).

And Le Roy rhetorically asks (Le Roy, 1960):

> Nous ne savons aucunement d'avance quelle sera la réponse, ni quel remaniement l'épreuve nous forcera de faire subir au système antérieur at au concept nouveau, quel aspect final ils prendront l'un et l'autre (p. 299).

(Note that in this case one interesting "remaniement" led to Kummer's theory of ideal numbers; on this point see (Ellison, 1978, pp. 195-200.))

Considerations like these support Mach's well-known idea of a structural analogy between experiments in physics and demonstrations in mathematics (e.g. see (Mach, 1976). Indeed, this seems to explain why in general complex numbers offer a typical example of circumstances where "the body of mathematical tools anticipated the physicist's needs" (Thom, 1982).

2. THE I-STORY

Keeping this in mind, let us come back to the crucial object studied by mathematicians who were building an arithmetic for various $\mathbf{Z}[D]$ or directly for $\mathbf{C}$: the quantity i, where $i^2 = -1$. To begin with, consider the following quotation from Heaviside's *Theory of Electromagnetism* (*ETM*) (Heaviside, 1899):

> It is not so long ago since mathematicians of the highest repute could not see the validity of investigations based upon the use of the algebraic imaginary. The results reached were, according to them, to be regarded as suggestive merely, and required proof by methods not involving the imaginary. (p. 459)

Heaviside remarks that in a research of this kind, strict Euclideanism represents an obstacle.[1] To those critics who note that "the rigorous logic of

[1] "The reader who may think that mathematics is all found out, and can be put in a cut-and-dried from like Euclid, in proposition and corollaries, is very much mistaken; and if he expects a similar systematic exposition here he will be disappointed. The virtues of the academical system of rigorous mathematical training are well known. But it has its faults. As very serious one (perhaps a necessary one) is that it checks instead of stimulating any originality student may possess, by keeping him in regular grooves. Outsiders may find that there are other grooves just as good, and perhaps a great deal better, for their purposes. Now, as my grooves are not the conventional ones, there is no need for any formal treatment. Such would be quite improper for our purpose, and would not be favourable to rapid acquisition and comprehension. For it is in mathematics just as in the real world; you must observe and experiment to find out the go of it. All experimentation is deductive work in a sense, only it is done by trial and error, followed by new deductions

the matter is not plain", Heaviside replies (Heaviside, 1899): "Well, what of that? Shall I refuse my dinner because I do not fully understand of the process of digestion?" (p. 9).

Quite correctly, Heaviside (1899) insists on the need for algebra to reach "a certain stage of development" before the imaginary "turns up":

> It was exceptional, however, and unintelligible, and therefore to be evaded, if possible. But it would not submit to be ignored. It demanded consideration, and has since received it. The algebra of real quantity is now a specialisation of the algebra of the complex quantity, say $a + bi$, and great extensions of mathematical knowledge have arisen out of the investigation of this once impossible and non-existent quantity. It may be questioned whether it is entitled to be called a quantity, but there is no question as to its usefulness, and the algebra of real quantity would be imperfect without it. (pp. 457-458)

As has recently been suggested (Stillwell, 1989), the quantity i seemed unintelligible because "a square of negative area did not exist in geometry" (p. 189). Appeal to history is here fundamental. The same historian pinpoints (Stillwell, 1989):

> The usual way to introduce complex numbers in a mathematical course is to point out that they are needed to solve certain quadratic equations, such as equation $x^2 + 1 = 0$. However, this did not happen when quadratic equations first appeared, since at that time there was no *need* for all quadratic equations to have solutions. Many quadratic equations are implicit in Greek geometry, as one would expect when circles, parabolas, and the like, are being investigated, but one does not demand that every geometric problem have a solution. If one ask whether a particular circle and line intersect, say, then the answer can be yes or no. If yes, the quadratic equation for the intersection has a solution; if no, it has no solution. An "imaginary solution" is uncalled in this context. (p. 189)

Indeed, the origin of i as a "solution" of the equation $x^2 + 1 = 0$ is a myth (Giusti, 1999). The context for the imaginary quantity was the solution of the

---

and changes of direction to suit circumstances. Only afterwards, when the go of it is known, is any formal exposition possible. Nothing could be more fatal to progress than to make fixed rules and conventions at the beginning, and then go by mere deduction. You would be fettered by your own conventions, and be in the same fix as the House of Commons with respect to the despatch of business, stopped by its own rules" (Heaviside, 1899, pp. 32-33). On the limits of the Euclidean approach see also (Lakatos, 1976a, pp. 205-207).

cubic equation in the heroic age of the Italian algebra. In fact, the del Ferro-Tartaglia-Cardano solution of the cubic equation $y^3 = py + q$ is

$$y = \sqrt[3]{\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3}} + \sqrt[3]{\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3}} \quad [\ldots].$$

The formula involves complex numbers when $\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3 < 0$.

However, it is not possible to dismiss this as a case with no solution, *because a cubic always has at least one real root* (since $y^3 - py - q$ is positive for sufficiently large positive y and negative for sufficiently large negative y).

Thus the Cardano formula raises the problem of reconciling a real value, found by inspection, say, with an expression of the form (Stillwell, 1989, p. 189):

$$\sqrt[3]{a + b\sqrt{-1}} + \sqrt[3]{a - b\sqrt{-1}} \ .$$

The first work to take complex numbers seriously was not Cardano's *Ars Magna* (1545) (in spite of the phrase "Cardano's formula"), but Rafael Bombelli's *Algebra* (1572). We will not attempt a detailed historical discussion of the solutions to this particular paradox of the cubic equation. For us, obviously, the solution is connected with the nature of *i* and the geometrical explanation of the meaning of this symbol in the Wessel-Argand-Gauss geometrical interpretation (Kline; 1972; van der Waerden, 1985, 178). But this interpretation came centuries after Cardano's formula and the algebraic approach sketched in Bombelli's work! Moreover, the turning point occurred when Descartes, in his *Geometry*, merged the problem of the nature of square root of –1 with the more general problem of "demonstrating" the so-called fundamental theorem of algebra. As he wrote, every algebraic equation has many solutions as his degree, but these solutions "ne sont pas toujours reelles, mais quelquefois seulement imaginaires" (Descartes, 1637). Aptly, Giusti comments that (Giusti, 1999) "Descartes does not explain what these imaginary roots are, and we have to intend literaliter this adjective *imaginary*" (p. 90); see also (van der Waerden 1985, pp. 72-75).

Be that as it may, the general development of algebra needed the consideration of numbers like $a + b\ \sqrt{-1}$, as Heaviside pointed out. Today, we can say that (Stillwell, 1989)

> at the beginning of their history, complex numbers $a + b\sqrt{-1}$ were considered to be "impossible numbers", tolerated only in a limited algebraic domain because they seemed useful in the solutions of cubic equations. But their significance turned out to be geometric and ultimately led to the unification of algebraic functions with conformal mapping, potential theory, and another "impossible" field, non Euclidean geometry. This resolution of the paradox of $\sqrt{-1}$ was so powerful, unexpected, and beautiful that only the word "miracle" seems adequate to describe it. (p. 188)

This "miracle" is more astounding than the description of the *i*-story offered by Heaviside would suggest. However, Heaviside's account discloses an interesting pattern in the growth of mathematics: namely, the transition from intuition to geometrical rigour *via* a process guided by the reliance on the power of algebra, tested by some kind of "mathematical experiments". Even more significantly, he draws a comparison between Imaginary Quantity and his Operational Calculus, in particular with the so-called fractional differentiation (Heaviside, 1899):

> Now just as the imaginary first presented itself in algebra as unintelligible anomaly, so does fractional differentiation turn up in physical mathematics. It seems meaningless, and that suggests its avoidance in favour of more roundabout but understandable methods. But it refuses to be ignored. Starting from the ideas associated with complete differentiations, we come in practice quite naturally to fractional ones and combinations. This occurs when we known unique solutions to exist, and asserts the necessity of a proper development of the subject. Besides, as the imaginary was the source of a large branch of mathematics, so I think must be with generalised analysis and series. Ordinary analysis is a specialised form of it. There is a universe of mathematics lying in between the complete differentiations and integrations. The bulk of it may not be useful, when found, to a physical mathematician. The same can be said of the imaginary lore. (pp. 459-460)

We claim that an analogous pattern can be found in the Operational Calculus or in what we call the δ-story.

3. THE DRIVING FORCE OF "ALGEBRAICAL" IMAGINATION. THE δ-STORY

It is well known that Heaviside's main contribution to science was his development and reformulation of Maxwell's Electrodynamics.[2] It was in this context that his mathematical ideas concerning Vector Analysis and Operational Calculus arose. In both fields, Heaviside was a great *dissenter* with respect to the scientific community of his time. In what follows, we shall focus just on the Operational Calculus. In his classic article on Heaviside, sir Edmund Witthaker writes (Whittaker, 1928/1929):

> We should now (1928) place the Operational Calculus with Poincaré's discovery of automorphic functions and Ricci's discovery of the Tensor Calculus as the three most important mathematical advances of the last quarter of the nineteenth century. Applications, extensions and justifications of it constitute a considerable part of the mathematical activity of to day. (p. 216)

The same source emphasizes Heaviside's discomfort caused by criticism from Cambridge mathematicians (Witthaker, 1928/1929, pp. 211-216). In hindsight, however, we can say that it was precisely his experimental conception of mathematics, so despised by his purist critics, to lead him to the definition of operational methods and to the intuition of what would later be known as Dirac's δ-function.

In the rest of this section, we are going to offer a reconstruction of Heaviside's procedure with respect to some physical issues discussed in his *EMT*. Along the lines of (Lützen, 1979) and (Petrova, 1987) (see also (Struppa, 1983; Guicciardini, 1993)), though in a somewhat different way, we shall distinguish four steps in Heaviside's procedure:

a) operational solution
b) algebrization
c) fractional differentiation
d) impulsive function

(a)Operational solution

In *EMT* §§ 238-242, Heaviside considers a semi-infinite cable and a network with resistance operator Z in sequence, operated upon by an electromotive force E. Putting aside the self-induction in the cable, he finds that the potential $V(x, t)$ and the current $C(x, t)$ are connected by the equations:

[2] On Heaviside's life and work see (Süsskind, 1972; Nahin, 1988; Lynch 1991).

$$-\frac{\partial C}{\partial x}=SpV\,, \quad -\frac{\partial V}{\partial x}=RC \tag{1}$$

where S is the permittance, R the resistance per unit length, and p the Heaviside's notation for the differential operator $\frac{\partial}{\partial t}$. From Eqs.(1) Heaviside derives the "characteristic" or operational equation

$$\frac{\partial^2 V}{\partial^2 x}=RSpV=q^2V \tag{2}$$

where q is defined by $q^2=RSp$. If we treat q as a constant, the operational solution of Eq.(2) would be

$$V(x,t)=Ae^{qx}+Be^{-qx} \tag{3}$$

where A and B are arbitrary functions of t. Yet, A and B are determined from the boundary conditions at $x = 0$ and $x = \infty$ yielding

$$V(x,t)=V_o e^{-qx} \tag{4}$$

where V_0 is the impressed electro-motive force at the end ($x = 0$). By Eq.(4), and the second equation of Eqs.(1), we get

$$C(x,t)=\frac{q}{R}V=\frac{q}{R}e^{-qx}V_0=C_0e^{-qx} \tag{5}$$

where C_0 is the current at the end of the cable ($x = 0$). We have also

$$C_0=\frac{q}{R}V_0=\left(\frac{Sp}{R}\right)^{\frac{1}{2}}V_0 \tag{6}$$

and similarly,

$$V_0 = \frac{R}{q} C_0 = \left(\frac{R}{Sp} \right)^{\frac{1}{2}} C_0 \tag{7}$$

Thus, Heaviside can write that (Heaviside, 1899) "the resistance operator is $\left(\frac{R}{Sp} \right)^{\frac{1}{2}}$" (p. 34).

(b)Algebrization

Now, if we put the resistance operator Z between the cable and the earth with the impressed voltage acting, we have

$$C_0 = \frac{E}{1 + Z \left(\frac{R}{Sp} \right)^{\frac{1}{2}}} \tag{8}$$

to express the current through Z and entering the cable. As he explains, "this is because the operators are additive like resistances" (Heaviside, 1899, p. 37). Also, we have

$V_0 = \left(\frac{R}{Sp} \right)^{\frac{1}{2}} C_0$ as before; consequently, by Eq.(8) we get

$$V_0 = \frac{E}{1 + Z \left(\frac{Sp}{R} \right)^{\frac{1}{2}}} \tag{9}$$

The latter expresses V_0, the potential at the beginning of the cable, in terms of E. If we suppose that Z is a "mere resistance" r and that the impressed force E is "constant after $t = 0$, having previously been zero" (this is the fundamental hypothesis on the *physical* interpretation of the problem; this hypothesis can be translated in mathematical language simply writing $E = H(t)$, where $H(t)$ is the so called "Heaviside's function"), we can

"algebrize" Eq.(9), i.e. "convert [it] to algebraical form" (Heaviside, 1899, p. 37), by expanding in ascending powers of p:

$$V_0 = \left\{1 - r\left(\frac{Sp}{R}\right)^{\frac{1}{2}} + r^2\left(\frac{Sp}{R}\right) - r^3\left(\frac{Sp}{R}\right)^{\frac{3}{2}} + \ldots\right\}E . \qquad (10)$$

For integral values of n, Heaviside puts $p^n E = 0$ and obtains

$$V_0 = \left\{1 - r\left(1 + \frac{r^2 Sp}{R} + \frac{r^4 S^2 p^2}{R^2} + \ldots\right)\left(\frac{Sp}{R}\right)^{\frac{1}{2}}\right\}E . \qquad (11)$$

(c)Fractional differentiation

We come now to the step of fractional differentiation. Whittaker remarks (Whittaker, 1928/1929):

> This is an old subject: Leibniz considered it in 1695 and Euler in 1729: and indeed it was in order to generalize the equations
>
> $\frac{d^n\left(x^k\right)}{dx^n} = k(k-1)(k-2)\ldots(k-n+1)x^{k-n}$ to fractional values of n that Euler invented the the Gamma-Function (p. 213).

Since Leibniz, Johann Bernoulli and Euler times there has been an almost continuous succession of papers about fractional derivates (Ross, 1977); but, as Whittaker says, "Heaviside seems to have known nothing of them beyond a reference of few lines in Thomson and in Tait's *Natural Philosophy*: but he carried the subject on original lines further, in some directions, than any of his predecessors" (p. 213).

Heaviside's awareness of the relevance of the problem is striking. In *EMT* (§ 225) he asserts that "*physical problems lead to improved mathematical methods*". As we proceed in extending the electrical theory, "so it is in mathematics. The fundamental notions are so simple that one might expect that unlimited developments could be made without ever coming to anything unintelligible. But we do, and in various direction" (p. 8). A typical example, as we have seen, is represented by complex numbers. But (Heaviside, 1899)

> there are much more obscure and ill understood questions, such as the meaning and true manipulation of divergent series, and of fractional differentiations or integrations, and connected matters. It is customary to keep to convergent series and whole differentiations and regard divergent series and fractional differentiations as meaningless and practically useless, or even to ignore the altogether, as if they did not exist. The latter is the usual attitude of moderate and practical mathematicians, for obvious reasons. If they can be ignored, why trouble about them at all? But when these things turn up in the mathematics of physics the physicist is bound to consider them, and make the best use of them that he can. I am thinking more particularly here of the solution of the differential equations to which physicist are led by quasi-algebraical processes. [...] I suppose all workers in mathematical physics have noticed how the mathematics seems made for the physics, the latter suggesting the former, and that practical ways of working arise naturally. This is really the case with resistance operators. It is a fact that their use frequently effects great simplifications, and the avoidance of complicated evaluations of definite integrals. But then the rigorous logic of the matter is not plain! Well, what of that? Shall I refuse my dinner because I do not fully understand the process of digestion? No, not if I am satisfied with the result. Now a physicist may in like manner employ unrigorous processes with satisfaction and usefulness if he, by the application of tests, satisfies himself of the accuracy of his results. At the same time he may be fully aware of his want of infallibility, and that his investigations are largely of an experimental character, and may be repellent to unsympathetically constituted mathematicians accustomed to a different kind of work. (pp. 8-10)

In order to handle fractional differentiation, Heaviside deduces in a "purely experimental way" the "fundamental formula" (Heaviside, 1899)

$$p^{\frac{1}{2}}H(t) = (\pi t)^{-\frac{1}{2}} \tag{12}$$

where $H(t)$ "means that function of time which is zero before and unity after $t = 0$" (p. 36). We find here the function $H(t)$, defined as $H(t) = 1$, for $t \geq 0$; $H(t) = 0$, for $t < 0$.

From Eqs.(11) and (12), it follows that

$$V_0 = E - Er\left(\frac{S}{R\pi t}\right)^{\frac{1}{2}}\left\{1 - \frac{r^2 S}{2Rt} + 1\times 3\left(\frac{r^2 S}{2Rt}\right)^2 - \ldots\right\} \tag{13}$$

As Heaviside comments (Heaviside, 1899)

> when t is big enough, the only significant term is e, the final value. When t is smaller, the next becomes significant. When smaller still another term requires to be counted, and so on. But we must never pass beyond the smallest term in the series. As t decreases, the smallest term moves to the left. As it comes near the beginning of the series, the accuracy of calculation becomes somewhat impaired. When it reaches the first t term, so that the initial convergence has wholly disappeared, then we can only roughly guess the value of the series. So Eq.(13) is unsuitable when t is small enough to make the initial convergence be insufficient (p. 38).

However, “every bane has its antidote”, and amateur botanists know that “the antidote is to be found near the bane” (pp. 38-39).

In our case, the antidote is got by algebrizing Eq.(9) in “a different way”, i.e. by expanding the expression in Eq.(9) in descending powers of p

$$V_0 = \left\{\frac{R}{r^2 Sp} + \left(\frac{R}{r^2 Sp}\right)^2 + \ldots\right\}\left(\frac{r^2 Sp}{R}\right)^{\frac{1}{2}} E - \left\{\frac{R}{r^2 Sp} + \left(\frac{R}{r^2 Sp}\right)^2 + \ldots\right\} E \tag{14}$$

and by applying Eq.(12), to obtain

$$V_0 = 2E\left(\frac{Rt}{r^2 S\pi}\right)^{\frac{1}{2}}\left\{1 + \frac{2Rt}{3r^2 S} + \frac{1}{3\times 5}\left(\frac{2Rt}{r^2 S}\right)^2 + \ldots\right\} - E\left(\exp\frac{Rt}{r^2 S} - 1\right) \tag{15}$$

We see now that “we can calculate V_0 conveniently when t is small”. But, as adds Heaviside (Heaviside, 1899), “(15) is bad when t is big. Then we may consider (15) the bane, and (13) the antidote. They are complementary, though not mutually destructive” (p. 39).

(d)The impulsive function

As we saw in (c), Heaviside puts $p^n E = p^n H(t) = 0$. However, as Lützen remarks (Lützen, 1979), "in other connections [...] he often showed a deeper understanding of $p^n H(t)$ considering it the 'function' similar to what we denote by δ^{n-1}" (p. 174). For instance, in EMT § 249 (Heaviside, 1899) he considers the "interesting and instructive case" which arises "when the impressed force at the beginning of the cable, inserted between it and earth, is variable whit the time in a certain way" (p. 54). Let the impressed force be given by

$$E = Q\left(\frac{R}{S\pi t}\right)^{\frac{1}{2}} \tag{16}$$

where Q is a constant charge. For $t < 0$ the cable is to be understood uncharged. Obviously, the potential V_0 is raised to the value E, i.e.

$$V_0 = Q\left(\frac{R}{S\pi t}\right)^{\frac{1}{2}}. \tag{17}$$

From the definition of q^2 and from (12), it is easy to obtain

$$V_0 = \frac{qQ}{S} \tag{18}$$

Now we can find the current entering the cable due to the impressed force. By Eq.(5), this is

$$C_0 = \frac{q}{R}V_0 = \frac{q^2}{RS}Q = pQ \tag{19}$$

where the second equation arises by Eq.(18), and the third by the definition of q^2. Heaviside is thus led to conclude that (Heaviside),

> since Q is constant for any finite value of time, the result is zero. That is, there is no current entering the cable under the action of the continuously-present impressed force at any finite value of the time (pp. 54-55).

Even more important is his remark (Heaviside, 1899):

> Is it nonsense? Is it an absurd result indicating the untrustworthy nature of the operational mathematics, or at least indicative of some modification of treatment being desirable? Not at all. […] We have to note that if Q is any function of the time, then pQ is its rate of increase. If, then, as in the present case, Q is zero before and constant after $t = 0$, pQ is zero except when $t = 0$. It is then infinite. But its total amount is Q. That it to say, $p\mathbf{1}$ means a function of t which is wholly concentrated at the moment $t = 0$, of total amount 1. It is an impulsive function, so to speak. The idea of an impulse is well known in mechanics, and it is essentially the same here. Unlike the function $p^{1/2}$ $\mathbf{1}$ [$\mathbf{1} = H(t)$], the function $p\mathbf{1}$ does not involve appeal either to experiment or to generalized differentiation, but involves only the ordinary ideas of differentiation and integration pushed to their limit. Our result $C_0=pQ$ therefore means that an impulsive current, that is a charge, is generated by the impressed force at the first moment of its application; that the amount of the charge is Q, and that there is no subsequent current. It is the same as saying that the charge Q is instantaneously given to the cable at its beginning, which charge then spreads itself without loss anything. (pp. 54-55)

It is obvious for us to find in this description a very interesting "mathematical object", i.e. Dirac's "δ-function" (Dirac, 1947). We also know the "happy end" of the story, i.e. the rigorous reformulation of Operational Calculus (Lützen, 1979), for example in the context of Laurent Schwartz's distribution theory (Schwartz, 1966).[3] Moreover, the "discovery" of "δ-function" by Heaviside in *EMT* is not the *first* discovery of it (Lützen, 1982). Yet, Heaviside's presentation is perhaps the most striking formulation before Dirac's, and it shades light on the relevance of the physical and/or technological context.

4. SOME CONCLUSIVE REMARKS

In spite of their different origins, the δ-story parallels the *i*-story (as Heaviside himself remarks). As mentioned before, the Cambridge mathematicians rejected Heaviside's procedures, i.e. "demonstrations" like sequence (1)-(19); they were less than satisfied by Heaviside's justification

[3] As Synowiec remarks (Synowiec, 1983), not only did Schwartz write (Schwartz, 1945) the "first systematic paper" on the theory of distributions, which "already contained most of the basic ideas, but he also wrote expository papers on distributions for electrical engineers (1948)".

of it, according to which "the use of operators frequently effects great simplifications, and the avoidance of complicated evaluations of definite integrals" (Heaviside, 1899, p. 9). (For the refusal from "pure" mathematicians see also (Hunt, 1991; Guicciardini 1993))

Yet, as Lützen (1982, p. 120) notes, Heaviside's polemic was not only just directed against Cambridge mathematicians but also against some *engineers* and technological people who might on the one hand be inclined to accept Heaviside's appeal to physical intuition, but were suspicious of his algebraic imagination on the other. Paraphrasing Lakatos, we can say that this fact is not just an historical oddity: it is also a sign of the *quasi-empirical* character of Heaviside's procedures. His apology on "the derivative of the *H*-function" reveals that in Heaviside's mathematical practice (and also in his idea of science), mathematics grows moving from physics (this is its *empirical* character), but proceeds by using "algebraical" tools in a novel way, stretching (in Lakatos' sense)– or "pushing to their limit" (in Heaviside's words)– standard concepts for new applications, and eventually testing the whole thing with mathematical experiments (this is its *quasi-empirical* character, and the emphasis on "quasi" now is crucial). He writes (Heaviside, 1899):

> It may be remembered that I have insisted upon the definitess and fullness of meaning of an operational solution, and that it contains within itself not only the full statement of the problem, but also its solution. No external aid is therefore required to algebrise it fully; no assumption, for instance, of a special type of solution, and that the solution is the sum of a number of that type, with subsequent determination of the constants required to complete the matter. The work of satisfying the imposed conditions has been done already. The conversion to algebraical or quantitative form may be easy or hard, self-evident or very obscure. But in any case it is possible, by the prior construction of the operational solution. Thus, the conversion furnishes a distinct subject of study which is of great practical value from the physical standpoint. As regards finding out how to effect the conversion, that is a matter principally of observation and experiment, and is in a great measure independent of logical demonstrations. It is the How, rather than the Why, with which we are mainly concerned in the first place; though, of course, parts of the Why cannot fail to be perceived in the course of examination of the How. A complete logical understanding of the subject implies the existence of a full theory to account for why certain ways of working are successful, and others not. It is important to note that it is just the same in the mathematical research into unknown regions as in experimental physical research. Observations of facts and experiments come first, with merely

> tentative suggestions of theory. As the subject opens out, so does the theory improve. But it can only become logical when the subject is very well known indeed, and even then it is bound to be only imperfectly logical, for the reasons mentioned at the beginning of this volume. I feel inclined to be rather emphatic on the matter of the use of experiment in mathematics, even without proper understanding. For there is an idea widely prevalent [...] that in mathematics, unless you follow regular paths, you do not prove anything; and that you are bound to fully understand and rigorously prove everything as you go along. This is a most pernicious doctrine, when applied to imperfectly explored regions. Does anybody fully understand anything? (pp. 122-123)

It is important to stress that "full understanding" may be impossible (as Lakatos says: in theory, research never ends). Nevertheless, this kind of quasi-empiricism in mathematics is a good tool for understanding, i.e. representing and intervening in the constitution of a mathematical object. For example, consider that the final (*for us*) part of the δ-story is not only some standard distribution theories (Sobolev spaces, Schwartz's theory, etc.), but reformulations of "the derivative of the *H*-function" in the context of Robinson's non-standard analysis (e.g. see (Robinson, 1966)) or another version of it.

Well, as in the case of Imaginary Quantity, in this case as well the starting point was a typical method of quasi-empirical mathematics offering "demonstrations" – and only *after* mathematicians were able to find rigorous "proofs".[4]

Still, even here we can talk of mathematical facts (in Le Roy's sense), namely the required properties of the "derivative of the *H* function".[5]

[4] *Demonstration* (from Latin *demonstratio*) means (i) "the action of showing forth or exhibiting", (ii) "the action or process of [...] making evident by reasoning", (iii) the "explanation of specimens and practical operations", (iv) and also "a public manifestation". While *proof* (late Latin *proba*, old French *prouve* and Italian *prova* or *pruova*) means "evidence sufficient (or contributing) to establish a fact or produce belief in the certainty of something" (OED).

[5] "Four different definitions or characteristic properties were mentioned in the literature before 1945:

$$\delta(x) = \frac{\partial}{\partial x} H(x);$$

$$\delta(x) = \lim_{n\to\infty} f_n(x) \text{ or } \delta = \sum_{n=0}^{\infty} f_n \text{ for suitable functions } f_n;$$

Moreover, the *results* obtained by Heaviside's "experimental method" are mathematically correct! Subsequent "rigorizations" have explained the reason why, connecting the Operational Calculus with other topics in mathematics and physics which form the context of the "prehistory" (see (Lützen, 1982, 163-165) of the rigorous theory of distributions (see (Schwartz, 1966); for a more general context see (Dieudonné, 1970, 1975). This is also the case for Dirac's δ in the non-standard analysis (e.g. see (Giorello, 1973).

Thus, the rigid opposition between discovery and invention is, with respect to mathematics, misleading; moreover, the picture of the growth of mathematics as "quasi-empirical" in Le Roy's or in Lakatos' sense is not so narrow as a simple step-by-step translation from mathematical language to physical language may suggest. If we look at the history of mathematical practice, both the *i*-story and the δ-story demonstrate that we have many layers of abstractions from scientific practice (physical or mathematical, for instance); for every layer, new objects obtained by abstraction are checked or tested by *results* or *needs* coming from former levels (*a contrario* evidence): think of the charge against Heaviside by engineers worried about *too much* "abstraction"! These engineers simply misunderstood the nature of Heaviside's "experimental" method, that is, the quasi-empirical pattern in the growth of mathematics.

If engineers in the 19th Century Britain feared that algebraic imagination could "pervert" their own practice, in the 20th Century some "pure" mathematicians maintained that mathematics "may be compared to a game – or rather an infinite variety of games" (Stone, 1961): two faces of the same picture! Yet, the story we have told shows that quasi-empiricism aims at reinstating into mathematics some "content". We insist that the prefix *quasi* is here the crucial term, because empiricism in Heaviside's sense needs a "whiff" of dialectics (Lakatos, 1976b; Motterlini, 2000; Lützen, 1982) in order to qualify the same idea of *experience* (consider l'*expérience mathématique* in Le Roy's sense).

$\delta(x) = 0$ for $x \neq 0$, and $\int_{-\infty}^{\infty} \delta(x) = 1$;

$\int_{-\infty}^{\infty} \delta(x-a) f(x) dx = f(a)$, or $\int_{-\infty}^{\infty} \delta(x) f(x) dx = f(0)$". (Lützen, 1982, p. 130)

In "experimental mathematics" (in Heaviside's terminology) we find a starting point for considering mathematical experience as a continuum which has as its poles motives coming from physics, technology, and so on, on the one hand, and "mathematical facts" apparently belonging in the domain of "pure mathematics", on the other.

Now, at least in some cases (e.g. distribution theory, functional analysis, but also Calculus, variation theory, differential equations, and so on), highly sophisticated mathematical methods give generality and soundness precisely to those original "experimental" methods that we have described. This move, moreover, seems to correspond in some aspects to the classical request of "geometrical rigour" in the *i*-story.

REFERENCES

Bombelli, R., 1966, *L'algebra*, Feltrinelli, Milano.

Boutroux, P., 1920, L'idéal scientifique des mathématiciens: dans l'antiquité et dans le temps modernes, Alcan, Paris, pp. 190-231.

Cardano, G., 1993, *Ars Magna or the Rules of Algebra*, translated and edited by T. R. Witmer, Dover, New York.

Cellucci, C., 2000, The growth of mathematical knowledge: An open world view, in *The Growth of Mathematical Knowledge*, E. Grosholz, H. Berger (ed), Kluwer, Dordrecht, 153-176.

Crowe, M. J., 1975, Ten 'Laws' concerning the patterns of change in the history of mathematics, *Historia Mathematica* **2**: 161-166.

Descartes, R., 1637/1987, *Discours de la méthode pour bien conduire sa raison, & chercher la vérité dans les sciences, plus La dioptrique, Les météores et La géométrie qui sont des essais de cette Méthode*, rist. anastatica, Conte, Lecce, p. 380.

Dieudonné, J., 1970, *Éléments d'analyse*, tome III, Gauthier-Villars, Paris, pp. 223-336.

Dieudonné, J., 1975, *Éléments d'analyse*, tome VI, Gauthier-Villars, Paris, pp. 131-190.

Dirac, P. A. M., 1947, *The Principles of Quantum Mechanics*, 3rd ed. Clarendon Press, Oxford.

Ellison, W.J. and F., 1978, Théorie des nombres, in J. Dieudonné (ed), *Abrégé d'histoire des mathématiques*, vol. I, Hermann, Paris, pp. 165-333.

Euclid, 1956, *The Thirteen Book of Euclid's Element*, vol. II, Dover, New York, p. 402.

Gillies, D., 2000, An Empiricist Philosophy of Mathematics and Its Implications for the History of Mathematics, in *The Growth of Mathematical Knowledge*, Grosholz E., Berger H. (ed), Kluwer, Dordrecht, 41-57.

Giorello, G., 1973, Una rappresentazione non-standard delle distribuzioni temperate e la trasformazione di Fourier, *Bollettino della Unione Matematica Italiana* **4**: 156-167.

Giorello, G., and Sinigaglia, C., 2001, *Pierre de Fermat. I sogni di un magistrato all'inizio della matematica moderna*, Le Scienze, Milano, pp. 90-93.

Giusti, E., 1999, *Ipotesi sulla natura degli oggetti matematici*, Bollati Boringhieri, Torino.

Guicciardini, N., 1993, Sulla variabilità dei criteri di rigore in matematica: il caso del calcolo operazionale di Oliver Heaviside, *Nuova Secondaria* **11**: 32-35.

Hadamard, J., 1949, *Psychology of Invention in the Field of Mathematics*, Princeton University Press, Princeton.

Heaviside, O., 1899, *Electromagnetic Theory*, Vol. II, The Electrician Printing and Publishing Company, London.

Heath, T. L., 1981, *A History of Greeks Mathematics*, Dover, New York, p. 401.

Hunt, B. J., 1991, Rigorous Discipline: Oliver Heaviside versus the Mathematicians, in *The Literary Structure of Scientific Argument. Historical Studies*, Peter Dear (ed), University of Pennsylvania, Philadelphia.

Kline, M., 1972, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, Oxford.

Lakatos, I., 1976a, A renaissance of empiricism in the recent philosophy of mathematics, *British Journal of Philosophy of Science* **27**: 201-233.

Lakatos, I., 1976b, *Proofs and Refutations. The Logic of Mathematical Discovery*, Cambridge University Press, Cambridge.

Le Roy, E., 1960, *La pensée mathématique pure*, Presses Universitaires de France, Paris, p. 304.

Lynch, A. C., 1991, The sources for a biography of Oliver Heaviside, *History of Technology*: **13**: 145-149.

Lützen, J., 1979, Heaviside's operational calculus and the attempts to rigorize it, *Archive for History of Exact Sciences* **21**: 161-200.

Lützen, J., 1982, *The Prehistory of the Theory of Distributions*, Springer, New York, Heidelberg, Berlin.

Mach, E., 1976, *Knowledge and Error. Sketches on the Psychology of Enquiry*, Reidel, Dordrecht, pp. 144-145.

Moore, D. H., 1971, *Heaviside Operational Calculus. An Elementary Foundations*, American Elsevier Publishing Company, New York.

Motterlini, M., 2000, *Lakatos. Scienza, matematica, storia*, Il Saggiatore, Milano, pp. 97-127.

Nahin, P. J., 1988, *Oliver Heaviside. Sage in Solitude*, IEEP Press, New York.

Petrova, S. S., 1987, Heaviside and the development of the symbolic calculus, *Archive for History of Exact Sciences* **37**: 1-23, pp. 14-17.

Ribenboim, P., 1979, *13 Lectures on Fermat's Last Theorem*, Springer, New York, pp. 35-82.

Robinson, A., 1966, *Nonstandard Analysis*, North-Holland, Amsterdam, pp. 133-145.

Ross, B., 1977, The development of fractional calculus, *Historia Mathematica* **4**: 75-89.

Schwartz, L., 1945, Généralisation de la notion de fonction, de dérivation, de transformation de Fourier et applications mathématiques et physiques, *Annales de l'Université de Grenoble* 21: 57-74.

Schwartz, L., 1948, Généralisation de la notion de fonction et de dérivation. Théorie des distributions, *Annales des Telecommunications* **3** : 135-140.

Schwartz, L., 1966, *Théorie des distributions*, Nouvelle éditions, Hermann, Paris.

Stillwell, J., 1989, *Mathematics and Its History*, Springer, New York.

Stone, M., 1961, The revolution in mathematics, *American Mathematical Monthly* **68**: 715-734.

Struppa, D., 1983, Cenni sulla nascita della teoria delle distribuzioni, *Quaderno S 18*, Dipartimento di Matematica "F. Enriques", Università di Milano.

Süsskind, C., 1972, Oliver Heaviside, in *Dictionary of Scientific Biography*, Charles Scribner's Sons, New York, 6 vol., 211-212.

Synowiec, J., 1983, Distributions: The evolution of a mathematical theory, *Historia Mathematica* **10**: 149-183, p. 149.

Thom, R., 1982, The Role of Mathematics in Present-Day Science, in *Logic, Methodology and Philosophy of Science IV*. Proceedings of the Sixth International Congress of Logic,

Methodology and Philosophy of Science, Hannover 1979, Cohen, L.J., Los, J., Pfeiffer, H., Podewski, K.-P. (ed), North-Holland, Amsterdam, New York, Oxford, 3-13, p. 4.

van der Waerden, B. L., 1985, *A History of Algebra. From al-Khwarizmi to Emmy Noether*, Springer, Berlin Heidelberg.

Whittaker, E., 1928/1929, Oliver Heaviside, *The Bulletin of the Calcutta Mathematical Society* **20**: 199-220; reprinted in (Moore, 1971).

QUANTUM PHYSICS AND MATHEMATICAL DEBATES CONCERNING THE PROBLEM OF THE ONTOLOGICAL PRIORITY BETWEEN CONTINUOUS QUANTITY AND DISCRETE QUANTITY

VICTOR GÓMEZ PIN
Universitat Autonoma de Barcelona, Barcelona, Spain

Abstract: In his book about the Categories (that is about the ultimate elements of classification and order), in the chapter concerning the quantity (IV, 20) Aristotle says that this concept recovers two kinds of modalities: the discrete quantity and the continuous quantity and he gives as examples the number for the first one; line, surface, solid, times and space for the second one. The main philosophical problem raised by this text is to determine which of the two modalities of the quantity has the ontological priority over the other (given two concepts A and B, we assume that A has ontological priority over B if every entity that possesses the quality B possesses necessarily the quality A). The problem is magnified by the fact that space, which in some part of Aristotle's *Physics* is mentioned not only as a category properly speaking but even as the main category whose power can be amazing, is in the evoked text of the Categories's Book reduced to expression of the continuum, and sharing this condition with time. In this matter the controversy is constant through the common history of Science and Philosophy.

In this paper we will recall the main points of projection of the controversy through the history of thought, from Zeno's aporias (and the mathematical attempts of solution) to the contemporary non standard analysis. To summarize: in order to display the ontological weight of quantum physics we will replace in its philosophical background the dramatic moment when Einstein suggested that Max Planck's theory was faraway of being merely an speculative mathematical construction, and that energy in nature actually comes in indivisible packets, instead of infinitely divisible streams. We will ask ourselves what different answers to the question have been brought forward by the ulterior developments of the discipline. In a second part of the

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 33–42.

paper we will try to establish the link between the problem raised up, the controversies about quantum non locality and the contemporary philosophical objections concerning the lack of rational explanation in the quantum theory, in spite of being largely successful at predicting the results of atomic processes. For, as the Newton's hypothesis *non fingo* displays, description and prevision does not necessarily means explanation.

Key words: continuous quantity; discrete quantity; ontological priority; quantum physics; locality; description versus explanation; Aristotle; Max Planck; Cantor; Einstein.

An electron can revolve in an orbit around the nucleus without losing energy, provided that the orbit... is a whole number of de Broglie wavelength in circumference.

Even if it has later been replaced by a more accurate model Bohr's picture (in 1913) of an electron's behaviour (completed by the de Broglie's idea of electron waves, idea that Bohr did not have, the whole set corresponding in fact to the ingenuous mental image that non specialist cultured people have of the atom) is well adjusted to the philosophical question that we would raise.

The problem is centred on the connected words whole and number. Of course at a certain level everyone knows what these expressions mean, and we are conscious that the scientist's work can hardly be subordinated to the philosophical controversies about the basic concepts of the mathematical background of the discipline.

Nevertheless, contemporary physics (quantum physics as well as relativity theory) has shown the impossibility of preserving the rigid division between scientific and philosophical work; and this is for two reasons:

a) Some of the prominent, and unanswered questions raised by quantum physics become necessary outstanding topics in the philosophy of science.

I have taken from a remarkable article of Nancy Cartwright (Cartwright, 1979) a provisional list of these topics:

- The nature of the time-energy uncertainty relations.
- Backward causation.
- Existence of photons.
- The nature of coherence.

b) The second and perhaps more important argument: nowadays when physicists have a look at the problems which would have arisen in the conceptual background of their disciplines by the very discoveries they make, they find exactly the same problems which are found in the classical content of ontology. In short: present physical knowledge tends intrinsically to become a reflexion about the categories that constitute the basis of the discipline itself and perhaps the basis of human knowledge... In this way

contemporary physics re-establishes a link with Greeks' physics, and particularly with Aristotle's physics steps: the description of the *physis* behaviour, the explanation of the recorded phenomena by displaying the causes that are immediately operating and finally, afterwards, -*meta*-research concerning the real explanation (looking after the last causes) which implies the evoked reflexion about the concepts presupposed by the previous work, and perhaps modified by the issues themselves (for example the backward causation as revolution in the concept of cause itself).

By reintegrating the third step the physics of today again becomes intrinsically reflexion after work, literally, *meta-ta-physika*.

Before coming to the main point of my purpose, I would like to recall that the giving up of the third step, the rupture of the intrinsic link between science and philosophy has a nitid birthday (even if it was announced by previous tendencies). In fact the official assumption of the divorce adopted the form of a new marriage: the repudiated philosophy as research of the last causes, and of a global explanation, was replaced by a philosophy "light" christened by its own promoter as experimental. Let us quote the text of Newton himself, the triumphal "hypothesis *non fingo*" in the scolia general of the third book of *Principia Mathematica*:

> I haven't succeeded in deducing from the phenomena the grounds of gravity and hypothesis *non fingo*, for everything that doesn't come from phenomena is a hypothesis and hypothesis can't be admitted into experimental philosophy. In this philosophy the propositions are extracted from the phenomena and next they are generalized by way of induction.

In fact Newton's hypothesis *non fingo* is not faraway from the concept of science that Roberto Bellarmino displayed in his famous letter of April 1512 to P. A. Foscarini, a friend of Galileo, in order to warn the latter about the danger of considering mathematical models operating from the point of view of the description of the phenomena as the real knowledge of these:

> ... Perche il dire che, supposto che la terra si muova e il sole stia fermo si salvano tute l'apparenze meglio che con gli eccentrici ed epicicli è benissimo detto e non ha pericolo nessuno; e questo basta al mathematico: ma volere affermare che realmente il sole stia nel centro del mondo... e cosa molto pericolosa non solo di irritare tutti i filosofi e theologi scholastici ma anche di nuocere alla Santa Fede con rendere false le Scritture Sante...

Describe and compute the phenomena "*basta al matematico*" whose work has nothing to do with the truth...

Analogously we could say: the phenomena follow a pattern that Newton's formula describes ... But these formulae have nothing to do with the cause of such behaviour.

So, experimental philosophy may replace classical natural philosophy. Leibniz was ready to object that the new theory was in fact a renouncement of the philosophy which intrinsically "cherche la raison et la divine sagesse qui la fournit"[6].

Moreover Leibniz pretends that experimental philosophy does not actually give up hypothesis. Simply the very fruitful hypothesis, these that would display reason (causal or not) of the phenomena are replaced by hypothesis that merely show some kind of correlation with the phenomena; hypothesis in fact asthenic (*hypothèses fainéants*, as Leibniz wrote) which would imply for science a complete rupture with the demands of rationality.

In comparison with this very hard appreciation of Leibniz, the observations concerning the disturbing paradoxes in some important aspects of quantum physics look almost like friendly encouragements to improve the general presentation of the theory. In fact the first people to keep their distance with the theory were the quantum-physicists themselves, which supposed a radical difference in relation to the Leibniz-Newton controversies.

In a letter of December 1926 Einstein claims that in spite of the fact that it describes the world with a level of accuracy without precedent in science, quantum mechanics (as a theory which would reduce God to a die player) "is not yet the real thing ... hardly brings us closer to the secret of the Old One" (Born and Einstein, 1971). But it was Einstein himself who 20 years before was compelled to consider the beam of light as an amount of discrete entities. An attitude nothing to do with Newton self indulgence concerning the lack of explanations (therefore the lack of meaning) in gravitational theory.

Einstein was confronted to a real contradiction between two scientific theories that he himself had helped to set up, and he could hardly be true to both of them. The choice had to be made, and Einstein's option became unequivocal in 1935 with the publication of Einstein's, Podolsky's and Rosen's famous paper (Einstein, Podolsky and Rosen, 1935)[7] concerning the

[6] 1978, Cherche la raison et la divine sagesse qui la fournit, Nouveaux Essais, T. V. Gerhardt, reed., Hildesheim N. Y. Olms, p. 39.

[7] At the symposium "New Developments on Fundamental Problems in Quantum Mechanics" celebrated in August 1995 in Oviedo, Spain, Professor Fine of Evanson has precised which had been the real role of Einstein in relation to the reduction of this paper. Fine said namely that Einstein had red the paper only after the first publication and even that he had

so called EPR effect, that is: a causal dependence between two phenomena which, according to the premises of quantum mechanics, should invalidate the thesis of the absolute speed of light, which was a basis of relativity theory.

It is not necessary to recall the Einstein Podolsky Rosen paper has to be completed by J. S. Bell's reflexion (Bell, 1964) since he demonstrated that the statistical laws of quantum theory are not compatible with both of the following assumptions: a) the universe has objective behaviour; b) every causation is necessarily local.

Certainly after 1964 number of purposes concerning this problem have been brought forward, but in substance the Bell's alternative remains: either Quantum Physics with just local causation but also no independent world, or Quantum Physics with causation necessarily "superluminal"[8].

We are here in presence of a controversy concerning concepts so important as *causa*, effect, temporality, movement and locality which form the nucleus of the traditional ontological problematic.

As N. Cartwright suggests, contemporary physics tends intrinsically to re-establish the link with the problems of philosophical tradition: the physicist, and especially the quantum physicist, without leaving his specific discipline, is sometimes confronted with questions which involves important discussions on ontology and epistemology.

But among the general issues on ontology one of the most controversial (at least from the point of view of a mathematical contemporary perspective) concerns the word that qualifies quantum physics itself: "quantity", *poson* which Aristotle tried, perhaps for the first times, to determine; this not through showing the specific differences with other categories (quality particularly) what should be impossible in Aristotle's conception of the ontological status of the first concepts but by displaying the internal division of the concept itself. Indeed, in his work known as *Categorie's Book* Aristotle writes (Aristotle, 1928):

> Quantity (*poson*) is either discrete (*diorismenon*) or continuous *(suneches*). Moreover, some quantities are such that each part of the

---

manifested some disagreement with the form (not of course with the content) and required some revisions before accepting to put his name in the second edition.

[8] For a general view of the issues of Bell's demonstration and its interpretations from Bernard d'Espagnat to Alain Aspect see (Woodhouse, 1992). The author pretends that, "barring current intuitive alternatives such as these stipulated by many-world theories", the "superluminal" theory should have the advantages of being: "a) consistent with QM; b) presupposed by the possibility of E. P. R. effects; c) derivable in part from the indivisibility of Planck's Quantum of actions, and, d) indirectly testable".

> whole has a relative position to the other parts, other have within them no such relation of part to part.
>
> Instances of discrete quantities are numbers and speech; of continuous: lines, surfaces, solids and, besides these time (*kronos*) and place (*pou*).
>
> In the case of the parts of a number there is no common boundary at which they join. For example: two fives make ten, but the two fives have no common boundary, but are separate [...] A line on the other hand, is a continuous quantity for it is possible to find a common boundary at which its parts join; in the case of the line, this common boundary is the point; in the case of the plane, it is the line [...].

This distinction between two concepts of quantity has to be completed by another, drawn in *Metaphysics* D13 (1020 a 7-12): "quantity (*poson*) is either plurality, multitude (*plethos*) or magnitude (*megethos*)". The first has to be understood as numerable (*arithméton*) while the second has to be understood as measurable (*metreton*). Moreover, while the first is divisible in parts that were discrete before the division and remain discrete afterwards, the second is divisible in parts that were continuous one to another and that after the division keep separated continuity.

Finally, there is a very important aspect: while the plurality (*plethos*) there is an absolute reference of enumeration, the unit, which is intrinsically indivisible[9] the magnitude (*megethos*) has not "natural" and absolute principle, but only an arbitrary and relative one.

In short: contrasting with the unit of enumeration, the unit of measure appears only as a result of "treating length as an atomic entity" (1052 b 32-33). This is because we do not consider the fact that the *metron* has a support that is intrinsically continuous, therefore intrinsically divisible.

We have in short:

Aristotle claims that the quantity as plurality does not have ontological objectivity, but it results merely from the capacity that the human mind has for making abstraction of the ontological reality. Meanwhile, for quantity as magnitude, the situation is not far from being the opposite, since the continuous is that to which we are confronted with, just when we stop in the work of making abstraction.

[9] The evoked texts have to be completed with Met. 10.1 1052b 22-23 ("The one is the first principle of number *qua* number") and following.

To summarize: while the unit of discrete quantity is a very (atomic) unit but ontologically is a vacuum, the unit of continuous quantity has great ontological weight but it is in fact a false (non atomic) unit.

The Aristotelian ontological priority of the continuum over the discrete quantity follows directly from these premises if we assume that given two concepts A and B, A has ontological priority over B if every entity that possesses the quality B also possesses the quality A. This definition of the ontological priority comes from René Thom who explains this idea more accurately in the following sentences (Thom, 1990a):

> L'être X est ontologiquement antérieur à l'être Y si et seulement si X peut recevoir naturellement Y comme prédicat, alors que X ne peut être que difficilement prédicat de Y... Ainsi il est lingüistiquement tout-à-fait acceptabe de parler d'une surface colorée alors qu'une couleur " "superficielle" ne pourrait s'employer que très métaphoriquement.

Before coming back to the main purpose of my talk let us recall some of René Thom's topological examples. The main idea is that a discrete entity is in fact a continuum that appears as discrete only by playing the role of attribute of a continuum of bigger dimensions (Thom, 1990b).

So if we consider a three-dimensional entity, its bidimensional surfaces are discrete accidents but in fact each of these is in itself a continuous support of lines as one-dimensional entities.

Moreover, the continuum remains locally continuous when considered as holding discrete attributes (like a body when determined by its dimension) meanwhile the entities playing the role of discrete attributes become locally continuous if they hold a continuous attribute.

The Aristotelian thought becomes clear when considered from Thom's contemporary perspective: the unit *arithmeton*, the foundation of discrete quantity, would never emerge, would never come to view, even merely as construction of the mind if the continuous was not there as intrinsic support.

It might be said that for Aristotle the indivisible produces only counting numbers (*arithmoi*): 1, 2, 3 and so on, which, as reduced to iteration of something intrinsically inexistent, are themselves merely abstractions of mind. On the other hand, the rational fractions (not *arithmoi*), which as results of the division of the unit interval [0, 1] would intrinsically pertain to the continuous, would so explain the structure of the substance itself. The latter is in fact a three-dimensional body, by essence intrinsically and infinitely divisible. Without continuous magnitude (*sunechen megethos*) not division and not rational fractions; a fortiori not irrational portions which, as

parcel of proportions theory, are not conceivable without reference to the *megethé*[10].

Therefore, the come back to Aristotle that implies René Thom's topological thought is completely in contradiction with the modern mathematics originated in the late nineteenth century with the works of Dedekind and Cantor, and which leads to a reduction of the continuous magnitude (*megethos*) to the discrete multitude (*plethos*).

In a work published in 1883, under the title of *Grundlagen allgenmeinen Manigfaltigkeitslehre*, Georg Cantor (Cantor, 1883/1932) displays his conviction that his theory of numbers carries out a revolution in both mathematical and philosophical reasoning, because it implies a modification in the concept of number itself and perhaps simply in the concept of quantity.

Of course the main point of this revolution is the legitimisation of magnitudes infinite *magnas*, which Leibniz repudiated[11], but a very important role is also played by the new mathematical conception of the continuum. Given an n-dimensional space, each point of this space is reduced to n-tuples defined from real numbers forming a complete ordered field which in fact find legitimacy in a reflexion submitted to the notion intrinsically discrete of cardinality. This last point becomes obvious if we consider the construction of real numbers as classes of equivalence of sequences of rational numbers; in such a way that every element of the complete field R becomes the limit of some denumerable, bounded and monotic sequence, either ascendent or descendent[12].

Well then:

In 1970 René Thom published a paper (Thom, 1970) that was a merciless condemnation of Cantor's views and proposed what René Thom himself called (some years later) "return to Aristotle". But the most explicit confrontation is the one that opposes René Thom to Dedekind's views.

[10] "It is instructive to note that in the primer of the theory of proportions that constitutes *Book 5* of Euclid, the definitions and theorems are all stated in terms of *megethé*." (White, 1992)

[11] At least in some texts: "Je leur tegmoirai que je ne croyais pas qu'il y eût des grandeurs vritablement infinies ni véritablement infinitesimals, ce que n'était que des fictions mais des fictions utiles pour abréger et pour parler universellement" (Letter to Dagincourt, 11 September 1916). The following sentence, which is in another register, is interesting too: "Mais comme M. le Marquis de l'Hôspital craignait que je ne trahisse la cause, ils me prièrent de n'en rien dire".

[12] Notice that the converse is also true: given a monotonic ascendent sequence of rational numbers, there exists a member of the complete ordered field R that is the limit of the sequence.

In his work 'Was sind and was sollen die Zahlen' Dedekind reduces the continuous line R to the completion of cuts (*schnits*) defined in rational numbers.

Evoking this text Thom writes (Thom, 1990a):

> Ici je voudrais m'attaquer à un mythe profondement ancré dans la mathématique contemporaine, à savoir que le continu s'engendre (voir se definit) à partir de la générativité de l'arithmétique, celle de la suitte des entiers naturels. Je fais, bien entendu, allusions à la construction de Dedekind.

But if we object to the arithmetical generativity of the continuum, we will be able to advance arguments backing the opposite hypothesis. Several attempts have been made before this of Professor Thom, which in fact has never been published in a very demonstrative way.

Let us simply recall that the link between contiguity and continuity has been reversed by Leibniz when he was confronted with the classical problem of the rupture of continuity, symbolized by the path from life to death. Leibniz's "solution" consisted of displaying the contiguity as a kind of rupture of topological continuity, an interpretation that would certainly appear scandalous to contemporary mathematical set theory[13].

First we had only a point C, and then ... we have two points A, B. Nevertheless these points are distinct but not distant; C has become a complex entity[14].

Of course contiguity as the rupture of continuity still does not mean discretion. But we are on the way... Leibniz's position shows that the problem of relationship between discrete quantity and continuous quantity is a nuclear point of the common history of philosophy and science. The history we have concentrated on is the debate Aristotle-Thom/Dedekind-Cantor. In this debate there is room for a third argument to be raised: that of non-standard analysis, but unfortunately there is not time to talk about it[15].

[13] As Enrico Giusti said (Giusti, 1986): "ciò che è privo di grandezza non è necessariamente privo di struttura".

[14] Nowadays, after Abraham Robinson's *Non Standard Analysis*, we could say that the structure C has only a lack of *grandezza* standard, for in the neighbour of C there exists hyperreal distances.

[15] Let us mention simply that the ontological weight of N. S. A. lies in the fact that for the first time since the problem occupied science and philosophy the infinitely small is legitimately introduced into Mathematics.

REFERENCES

Aristotle, 1928, *Categories's Book*, translated into English by E. M. Edghill, W. D. Ross, ed., Clarendon Press, Oxford.

Bell, J. S., 1964, On the Einstein Podolsky Rosen paradox, *Physics* **1**: 195-204.

Born, M., and Einstein, A., 1971, The Born-Einstein Letters, Walker, New York, p. 90.

Cantor, G., 1883/1932, Grundlagen allgenmeinen Manigfaltigkeitslehre, in: *Gesammelte Abhandlugen*, Springer Verlag, Berlin.

Cartwright, N., 1979, Philosophy in physics, in: *Current Research in Philosophy of Science*, P. D. Asquitt and H. E. Kyburg Jr., ed., Philosophical Association, East Landing, Michigan, pp. 382-384.

Giusti, E., 1986, Immagini del continuo nell'opera di Leibniz, Colloquio Leibniz-Gesselschaft, Roma, inedited.

Einstein, A., Podolsky, B., and Rosen, N., 1935, Can quantum mechanical description of physical reality be considered complete?, *Physical Review* **47**: 777-780.

Thom, R., 1970, Les mathématiques modernes, une erreur pédagogique et philosophique, in: *L'Age de la Science*, Villet Vol. III, Paris Dunod, pp. 225-242.

Thom, R., 1990a, L' anteriorité ontologique du continu sur le discret, in: *L'infini*, H. Sinacoeur, ed., Colloque de Cerisy.

Thom, R., 1990b, Un continu ne peut être discret que s'il apparaît comme un accident d'un autre continu de dimension supérieur (une droite dans le plan par exemple), in: *L'infini*, H. Sinacoeur, ed., Colloque de Cerisy.

White, M. J., 1992, *The Continuous and the Discrete. Ancient Physical Theories from a Contemporary Perspective*, Clarendon Press, Oxford, p. 185.

Woodhouse, M. B., 1992, Philosophy and frontier science, in: *New Horizons to the Philosophy of Science*, D. Lamb, ed., Avebury Series in Philosophy of Science, pp. 26-49.

JOHN VON NEUMANN ON MATHEMATICAL AND AXIOMATIC PHYSICS

MIKLÓS RÉDEI
Loránd Eötvös University, Budapest, Hungary

Abstract: The aim of this paper is to recall and analyse von Neumann's position on mathematical and axiomatic physics. It will be argued that von Neumann demanded much less mathematical rigor in physics than commonly thought and that he followed an opportunistically interpreted soft axiomatic method in physics. The notion of opportunistic soft axiomatization is illustrated by recalling his work on the mathematical foundations of quantum mechanics.

Key words: von Neumann; axiomatization; quantum mechanics.

1. TWO ATTITUDES TOWARDS MATHEMATICAL PRECISION AND AXIOMATIZATION IN PHYSICS

One of the key distinguishing features of physics has been since the dawn of the modern age that it is mathematical. By "being mathematical" I mean not only that it is quantitative in the sense that it gives a description of the physical world in numbers; rather, what is meant by being mathematical is that physics applies mathematical concepts above and beyond numbers, that it builds mathematical models of intricate structure by using sophisticated mathematical entities and procedures.

Merging of physics and mathematics on the non-numerical, conceptual level has never been unproblematic however: One just has to recall some of the famous conceptual-mathematical difficulties that accompanied the development of physics since the time of Galileo: the lack of a

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 43–54.

mathematically-logically acceptable calculus in Newton's physics, (so brilliantly pointed out and criticized by Berkeley (1948/1951)), the mathematically problematic status of the ergodic hypothesis in the work of Boltzmann on classical statistical mechanics (made clear and analysed by the Ehrenfests (Ehrenfests P. and Ehrenfests T., 1911)) and the mathematically unacceptable treatment of the eigenvalue problem of selfadjoint operators in quantum mechanics (pointed out and solved in full generality by von Neumann (1927a)) are well-known and much quoted classical examples of the inconveniences of the marriage between physics and mathematics.

There are two typical attitudes towards the conceptual difficulties arising from mathematical imprecision and sloppiness in physics: the easy-going and the concerned; accordingly, there are two attitudes towards the function of mathematical precision in physics: the sceptical and the reflective. The Sceptics claim that mathematical exactness is alien to and useless in physics. This view is explicitly formulated by R. Feynman, for instance (Feynman, 1965):

> The mathematical rigor of great precision is not very useful in physics. But one should not criticize the mathematicians on this score…They are doing their own job. (p. 56)

While Feynman's position might be typical in the physics community, the opposite, reflective position, according to which mathematical precision in physics is both needed and useful, has been successful enough to have led to a whole new discipline called mathematical physics. This field has become institutionalised in the 20th Century with a well-defined scientific community, with scholarly periodicals specializing in mathematical physics and with professional associations organizing the community of mathematical physicists.

The idea of mathematical physics has been intertwined in the 20th Century with another one that also is rooted deeply in mathematics: application of the axiomatic method in physics. Hilbert's sixth problem formulated this idea programmatically in 1900 (see (Wightman, 1976) for a review of Hilbert's sixth problem and its impact on the development of physics), and attempts have been made since to axiomatize all the basic physical theories.

Similarly to mathematical physics, axiomatic physics is not typically considered by physicists as especially useful. Weyl probably expresses the typical sentiment of the physics community's attitude towards the value of axiomatic physics when he writes (Weyl, 1944):

> The maze of experimental facts which the physicist has to take into account is too manifold, their expansion too fast, and their aspect and

> relative weight too changeable for the axiomatic method to find a firm enough foothold, except in the thoroughly consolidated parts of our physical knowledge. Men like Einstein or Niels Bohr grope their way in the dark toward their conceptions of general relativity or atomic structure by another type of experience and imagination than those of the mathematician, although mathematics is an essential ingredient. Thus Hilbert's vast plans in physics never matured. (p. 653)

John von Neumann is regarded by both the Skeptics and the Concerned as a typical mathematical physicist relying heavily on the axiomatic method. The aim of this paper is to describe von Neumann's position on mathematical and axiomatic physics. The analysis is motivated in part by what I take to be a somewhat curious situation: While mathematical physicists view his work as a paradigm example to be followed, and although even the Sceptics acknowledge that von Neumann's work is a great intellectual achievement, one hardly finds any detailed historical or philosophical analysis of his views on mathematical and axiomatic physics ((Halmos, 1973) and (Wightman, 1976) being exceptions). Lack of a careful study of von Neumann's views and of the method he actually followed in his work has led, I claim, to a somewhat distorted picture, showing him to demand far more mathematical rigor in physics than he actually did. I hope to be able to correct this one-sided picture of von Neumann.

2. VON NEUMANN ON MATHEMATICAL PHYSICS

The only explicit assertion by von Neumann I am aware of in which von Neumann discusses the nature of mathematical physics is in his letter to R. O. Fornaguera, the Spanish translator of von Neumann's book *Mathematical Foundation of Quantum Mechanics* (von Neumann, 1932). Von Neumann writes

> Your questions on the nature of mathematical physics and theoretical physics are interesting but a little difficult to answer with precision in my own mind. I have always drawn a somewhat vague line of demarcation between the two subjects, but it was really more a difference in distribution of emphasis. I think that in theoretical physics the main emphasis is on the connection with experimental physics and those methodological processes which lead to new theories and new formulations, whereas mathematical physics deals with the actual solution and mathematical execution of a theory which is assumed to be correct per se, or assumed to be correct for the sake of the discussion.

> In other words, I would say that theoretical physics deals rather with the formation and mathematical physics rather with the exploitation of physical theories. However, when a new theory has to be evaluated and compared with experience, both aspects mix. (quoted in (Redei, 2002), p. 242)

The position von Neumann takes in the above quotation concerning mathematical physics is a very moderate one: he does not see a neat separation of mathematics and theoretical physics and he takes the reflective nature of mathematical physics as its main characteristics – not mathematical exactness. "Reflective nature" means here that the immediate subject of mathematical physics is considered by him to be the physical *theory* rather than the physical world, the latter implicitly taken by von Neumann as the subject of theoretical physics. Investigating and "exploiting physical theories" is very much what philosophy of science (physics) typically does however, and I have argued elsewhere that this reflective nature of mathematical physics lends this discipline a philosophical character indeed (Redei, 2002). Von Neumann was very much aware of this feature of mathematical physics: he himself regarded his 1932 seminal work, "The mathematical foundations of quantum mechanics" (von Neumann, 1932) rather a conceptual-logical analysis than physics or mathematics: in characterizing the nature of this book, von Neumann points out in a letter to H. Cirker (von Neumann, 1949) that the real novelty and justification of the book is carrying out the very involved conceptual critique of the logical foundations of the relevant mathematical and physical discipline (such as theory of probability, thermodynamics, classical statistical mechanics and quantum mechanics.

3. VON NEUMANN ON THE AXIOMATIC METHOD

One can – and therefore has to – distinguish two different notions of "axiomatization" and "axiomatic theory" in von Neumann's works:

1. axiomatizing and axiomatic theory in the strict sense of formal (or syntactic) systems or languages (call this "formal axiomatization")
2. axiomatizing and axiomatic theory in the less formal sense in which it occurs in physics (following (Rédei and Stöltzner, forthcoming) - see also (Stöltzner, 2001) – let us call this "soft axiomatization")

Formal axiomatization is what von Neumann does in his work on axiomatic set theory (the topic of his PhD dissertation in 1926). This formal axiomatization is as it is understood today in the theory of formal languages

(syntactic systems). However, even in connection with formal axiomatization von Neumann takes a very sensible, only moderately formal position, making clear that there is always some intuitively given content or meaning behind the primitive concepts and the axioms in terms of which axiomatic set theory is formulated (von Neumann, 1927d):

> We begin with describing the system to be axiomatized and with giving the axioms. This will be followed by a brief clarification of the meaning of the symbols and axioms ... . It goes without saying that in axiomatic investigations as ours, expressions such as ‘meaning of a symbol’ or ‘meaning of an axiom’ should not be taken literally: these symbols and axioms do not have a meaning at all (in principle at least), they only represent (in more or less complete manner) certain concepts of the untenable ‘naive set theory’. Speaking of ‘meaning’ we always intend the meaning of the concepts taken from ‘naive set theory’. (p. 344) (translation from (Rédei and Stöltzner, forthcoming)

As opposed to formal axiomatization, soft axiomatization is a less well-defined, more intuitive and a structured concept. Its explicit formulation can be found in the 1926 joint paper by Hilbert, Nordheim and von Neumann on the foundations of quantum mechanics (Hilbert, Nordheim and von Neumann, 1926). This paper contains a lengthy passage on the axiomatic method in physics. The main idea is that a physical theory consists of three, sharply distinguishable parts:

1. physical axioms
2. analytic machinery (also called “formalism”)
3. physical interpretation

The physical axioms are supposed to be semi-formal requirements (postulates) formulated for certain physical quantities and relations among them. The basis of these postulates is our experience and observations; thus the basis of the axioms in physics is empirical, which is not necessarily the case in formal axiomatization: von Neumann points out that the fifth postulate in Euclid’s geometry is non-empirical.

The analytic machinery is a mathematical structure containing quantities that have the same relation among themselves as the relation between the physical quantities. *Ideally*, the physical axioms should be strong and rich enough to *determine* the analytic machinery *completely*. The physical interpretation connects then the elements of the analytic machinery and the physical axioms.

Here is the idea in the author’s words and specified for the case of quantum mechanics, where probability density for the distribution of values of physical quantities is taken as the basic, primitive concept (Hilbert, Nordheim and von Neumann, 1926):

> The way leading to this theory is the following: one formulates certain physical requirements concerning these probabilities, requirements that are plausible on the basis of our experiences and developments and which entail certain relations between these probabilities. Then one searches for a simple analytic machinery in which quantities appear that satisfy exactly these relations. This analytic machinery and the quantities occurring in it receive a physical interpretation on the basis of the physical requirements. The aim is to formulate the physical requirements in a way that is complete enough to determine the analytic machinery unambiguously. This way is then the way of axiomatizing, as this had been carried out in geometry for instance. The relations between geometric shapes such as point, line, plane are described by axioms, and then it is shown that these relations are satisfied by an analytic machinery namely linear equations. Thereby one can deduce geometric theorems from properties of the linear equations. (p. 105) (translation from (Rédei and Stöltzner, forthcoming).

Hilbert, Nordheim and von Neumann see clearly, however, that not even soft axiomatization is practiced in actual science. They point out that what happens is that one typically conjectures the analytic machinery *first* and *without* having formulated the physical axioms. It is only after the analytic, mathematical part is fixed that one gets insight into what the physical axioms should be. In their words (Hilbert, Nordheim and von Neumann, 1926):

> In physics the axiomatic procedure alluded to above is not followed closely, however; here and as a rule the way to set up a new theory is the following.
>
> One typically conjectures the analytic machinery before one has set up a complete system of axioms, and then one gets to setting up the basic physical relations only through the interpretation of the formalism. It is difficult to understand such a theory if these two things, the formalism and its physical interpretation, are not kept sharply apart. This separation should be performed here as clearly as possible although, corresponding to the current status of the theory, we do not want yet to establish a complete axiomatization. What however is uniquely determined, is the analytic machinery which – as a mathematical entity – cannot be altered. What can be modified – and is likely to be modified in the future – is the physical interpretation, which contains a certain freedom and arbitrariness. (p. 106) (translation from (Rédei and Stöltzner, forthcoming)

So, to the extent axiomatization is a method practiced in physics, it is only this soft axiomatization, and as Hilbert-Nordheim-Neumann point out, even this sort of axiomatization is typically practiced in a very opportunistic manner with many concessions to the given science's state of formalization.

It seems fair to say then that, according to the Hilbert-Nordheim-Neumann paper, axiomatization in physics is an *opportunistic soft axiomatization*, which seems such a soft notion indeed that one may wonder whether such a method should at all bear the name "axiomatization" and not be called simply "model building". Von Neumann would most likely not oppose such a terminology: in his later essays in which he addresses the issue of method in science he emphasizes precisely this feature of science (von Neumann, 1961):

> To begin, we must emphasize a statement which I am sure you have heard before, but which must be repeated again and again. It is that the sciences do not try to explain, they hardly ever try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of some verbal interpretations describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work – that is correctly to describe phenomena from a reasonably wide area.
>
> I will further limit myself to saying a few things about procedure and method which will illustrate the general character of method in science. Not only for the sake of argument but also because I really believe it, I shall defend the thesis that the method in question is primarily opportunistic – also that outside of the sciences, few people appreciate how utterly opportunistic it is. (p. 492)

To summarize: According to von Neumann, the method in the physical sciences is (and should be) a pragmatically interpreted opportunistic soft axiomatization.

4. CONCEPTUAL CLARITY IS MORE IMPORTANT THAN MATHEMATICAL PRECISION: VON NEUMANN'S WORK ON QUANTUM THEORY

It could in principle be that von Neumann's actual work in physics does not comply with the methodological prescription of pragmatically interpreted opportunistic soft axiomatization – but his work is in the spirit of

this methodological principle, and the aim of this section is to show this on the example of von Neumann's work on quantum mechanics.

Von Neumann's foundational work on quantum mechanics can be divided into two periods: the work between the years 1926-1932 and the post 1932 period. Von Neumann started working on quantum mechanics in 1926 while being an assistant of Hilbert in Göttingen. He published 3 papers (von Neumann, 1927a, 1927b, 1927c) in Göttingen and these papers served as the basis of his 1932 book (von Neumann, 1932) that summarizes what can be properly called the "Hilbert space quantum mechanics". This first period is the better known and it has been reviewed in (Jammer, 1974) for instance.

Less well known is that soon after von Neumann had finished his book, he started questioning the Hilbert space formalism, and by 1935-1936 he came to the conclusion that the Hilbert space formalism is *not* a suitable framework for quantum theory. Why? To understand von Neumann's position and in particular his abandoning the Hilbert space formalism one has to recall the core of the Hilbert space formalism as this was formulated in his 1932 book.

Von Neumann formulates only two, explicitly formulated physical axioms, both concern the nature of the expectation value of physical quantities in a statistical ensemble:

A Expectation value assignments $a \mapsto E(a)$ are linear:

$$E(\alpha a + \beta b + ...) = E(\alpha a) + E(\beta b)...$$

B Expectation value assignments are positive:

$$E(a) \geq 0 \qquad \text{if } a \text{ can take on only non} - \text{negative values}$$

These two postulates are informal and are based on empirical observations exactly in the sense in which the Hilbert-Nordheim-Neumann paper talks about physical axioms: The physical quantities a, b are left completely unspecified, and the two postulates spell out something that is a basic, empirically observable feature of expectation value assignments in a relative frequency interpreted probability theory.

The analytic machinery is the set of all selfadjoint operators on a Hilbert space, the third (**C**) and fourth (**D**) "postulates" specify the physical interpretation, the link between the physical quantities and the operators:

C If the operators $A, B...$ represent the physical quantities $a, b...$ then the operator $\alpha A+\beta B+...$ represents the physical quantity $\alpha a+\beta b...$

D If operator A represents the physical quantity a then the operator $f(A)$ represents the physical quantity $f(a)$.

It is rather obvious that the above axiomatization is indeed the sort of opportunistic soft axiomatization characterized in the Section 3: The opportunistic aspect of this soft axiomatization manifests in the fact that postulates **A** and **B** do *not* imply that the physical quantities need to be represented by the set of *all* linear operators on a Hilbert space. One has to, and von Neumann does indeed, *stipulate* that the physical quantities are represented by the formal machinery of linear operators on a Hilbert space.

From postulates **A+B+C+D** von Neumann deduces that every expectation value assignment is of the form

$$E(a)=Tr(UA) \tag{1}$$

with some statistical operator U (= positive, linear, not necessarily trace class!) and where the selfadjoint operator A represents physical quantity a.

Equation (1) is the heart of the whole theory, it contains all probability statements; specifically, according to von Neumann's interpretation, Eq. (1) yields the probabilities of quantum events:

$$p(P^A(d))=Tr(UP^A(d)) \tag{2}$$

where $P^A(d)$ is a spectral projection of some observable S with spectral measure P^A, the projection $P^A(d)$ representing the event that observable A takes its value in the set d of real numbers.

This is in a nutshell of von Neumann's approach to quantum mechanics in the years 1926-1932. Von Neumann realized however that his interpretation of the trace formula is beset with deep conceptual problems: in order to be able to interpret the probabilities $p(X)$ as relative frequencies (in von Mises' sense, where there is a *fixed* statistical ensemble in which one has to compute the probabilities) the probability assignment $X \mapsto p(X)$ needs to satisfy the following "subadditivity" property:

$$p(X)+p(Y)=p(X \wedge Y)+p(X \vee Y) \quad \text{for all projections } X,Y \tag{3}$$

where $\wedge$ and $\vee$ are the standard lattice operations between Hilbert space projections. But the subadditivity property is violated by every p defined by a *every* non-trivial statistical operator $U \neq I$; on the other hand, the "probabilities" given by the identity operator $U=I$ as statistical

operator are not finite, hence they *cannot* be interpreted as relative frequencies at all.

Von Neumann was struggling with this problem already in his second 1927 paper on the foundations of quantum mechanics (von Neumann, 1927b) and also in his book (von Neumann, 1932); and this conceptual problem was the main reason, I claim, why he lost his belief in the Hilbert space formalism by about 1935 (see (Rédei, 1996) for further details). Von Neumann's solution of this conceptual problem in 1935-1936 was that he suggested that the proper mathematical framework for quantum theory is the theory of type II_1 factor von Neumann algebras. He maintained this view as late as in his 1954 address on "Open Problems in Mathematics", which is his last word on quantum theory (see (Rédei, 1999) for details).

It is important to point out that von Neumann's preference of the theory of II_1 factors as the proper mathematical framework of quantum theory was *not* based on any mathematical imprecision in the Hilbert space formalism, nor was it motivated by any discovery of a new physical fact or phenomenon: it was motivated exclusively by informal, conceptual-philosophical difficulties related to the interpretation of probability in quantum theory. Thus one has to conclude that what drove von Neumann's research in physics was *not* his desire to have mathematically impeccable theories: It was more important for him to create theories that are conceptually sound. What better further proof of this claim can one have than the fact that von Neumann also realized that even taking the theory of II_1 factor von Neumann algebras as the proper mathematical framework for quantum theory does not solve the problem of how to interpret quantum probability, and in 1936 he finally abandoned the relative frequency view of quantum probabilities altogether (von Neumann, 1962):

> This view, the so-called 'frequency theory of probability' has been very brilliantly upheld and expounded by R. von Mises. This view, however, is not acceptable to us, at least not in the present 'logical' context. (p.196)

(See (Rédei, 1998, 1999, 2001) for further details of von Neumann's post 1932 views on quantum mechanics and quantum probability.)

5. SUMMARY

Contrary to what seems to be a common evaluation of von Neumann's position concerning the role of mathematical rigor in physics, von Neumann was very relaxed about mathematical precision in physics. A closer look at his views on the nature of the axiomatic method in mathematics (set theory)

and physics (quantum mechanics) show that he did not consider the axioms of set theory as a purely formal system and that he followed an opportunistically interpreted soft axiomatic method in quantum theory. Von Neumann's post 1932 work on quantum mechanics, and in particular his abandoning the Hilbert space formalism in favour of operator algebra theory, show that for him conceptual clarity and existence of an intuitively satisfactory interpretation of a physical theory were more important than its mathematical precision. Von Neumann was ready to abandon a physical theory, however clean mathematically, if it was conceptually problematic – like any truly deep thinker he was against any dogmatism in science.

ACKNOWLEDGEMENT

Work supported by OTKA (contract numbers: T 043642 and TS 040899).

REFERENCES

Berkeley, G., 1948/1951, The analyst, in: *The Works of George Berkeley Cloyne*, Vol. IV., Luce, A. A. and Jessop, T. E., (eds.), TH. Nelson and Sons LTD., London, pp. 65-102.

Ehrenfets, P., and Ehrenfest, T., 1911, Begriffliche Grundlagen der statistischen Auffassung in der Mechanik in: *Enzyclopädie der Mathematischen Wissenschaften* vol. 4. (Teubner, Leipzig, 1911). English translation: 1959, *The conceptual foundations of the statistical approach in mechanics*, Cornell University Press, Ithaca.

Feynman, R., 1965, *The Character of Physical Law*, MIT Press, Cambridge, Mass.

Halmos, P., 1973 : The legend of John von Neumann, *American Mathematical Monthly* **50**: 382-394.

Hilbert, D., Nordheim, L.,, and von Neumann, J., 1926, Über die Grundlagen der Quantenmechanik, in: 1962, *Collected Works Vol. 1 Logic, Theory of Sets and Quantum Mechanics*, A. H. Taub, ed., Pergamon Press, pp. 104-133.

Jammer, M., 1974, *Philosophy of Quantum Mechanics*, Wiley, New York.

Rédei, M., 1996, Why John von Neumann did not like the Hilbert space formalism of quantum mechanics (and what he liked instead), *Studies in the History and Philosophy of Modern Physics*: 493-510.

Rédei, M. 1998, *Quantum Logic in Algebraic Approach*, Kluwer Academic Publishers, Dordrecht.

Rédei, M., 1999, 'Unsolved problems in mathematics' J. von Neumann's address to the International Congress of Mathematicians, Amsterdam, September 2-9, 1954, *The Mathematical Intelligencer*: 7-12.

Rédei, M., 2001, John von Neumann's concept of quantum logic and quantum probability, in: *John von Neumann and the Foundations of Quantum Physics*, M. Rédei and M. Stöltzner, eds., Kluwer, Dordrecht, pp. 153-172.

Rédei, M., 2002, Mathematical physics and philosophy of physics (with special consideration of J. von Neumann's work), in: *History of Philosophy of Science. New Trends and*

Perspectives (Institute Vienna Circle Yearbook 2001), M. Heidelberger and F. Stadler, eds., Kluwer, Dordrecht, pp. 239-243.

Rédei, M., and Stöltzner, M., forthcoming, Soft axiomatism: von Neumann on method and von Neumann's method in the physical sciences, in: *Intuition and the Axiomatic Method*, E. Carson and R. Huber, eds., Western Ontario Series in Philosophy of Science.

Stöltzner, M., 2001, Opportunistic axiomatics: John von Neumann on the methodology of mathematical physics, in: *John von Neumann and the Foundations of Quantum Physics*, M. Rédei and M. Stöltzner, eds., Kluwer, Dordrecht, pp. 35-62.

von Neumann, J., 1927a, Mathematische Begründung der Quantenmechanik Göttinger Nachrichten, in: 1962, *Collected Works Vol. 1 Logic, Theory of Sets and Quantum Mechanics*, A. H. Taub, ed., Pergamon Press, pp. 151-207.

von Neumann, J., 1927b, Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik Göttinger Nachrichten, in: 1962, *Collected Works Vol. 1 Logic, Theory of Sets and Quantum Mechanics*, A. H. Taub, ed., Pergamon Press, pp. 208-235.

von Neumann, J., 1927c, Thermodynamik quantenmechanischer Gesamtheiten Göttinger Nachrichten, in: 1962, *Collected Works Vol. 1 Logic, Theory of Sets and Quantum Mechanics*, A. H. Taub, ed., Pergamon Press, pp. 236-254.

von Neumann, J. 1927d, Die Axiomatisierung der Mengenlehre, in: 1962, *Collected Works Vol. 1 Logic, Theory of Sets and Quantum Mechanics*, A. H. Taub, ed., Pergamon Press, pp. 339-422.

von Neumann, J., 1932, *Mathematische Grundlagen der Quantenmechanik*, Dover Publications, New York, first English translation: 1955, *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton.

von Neumann, J., 1949, Letter to H. Cirker (October 3, 1949), Library of Congress, Washington, D.C., unpublished. Forthcoming in: *John von Neumann's Selected Letters*, M. Rédei, ed.

von Neumann, J., 1961, Method in the physical sciences, in: *Collected Works Vol. VI, Theory of Games, Astrophysics, Hydrodynamics and Meteorology*, A. H. Taub, ed., Pergamon Press, pp. 491-498.

von Neumann, J., 1962, Quantum logics (strict- and probability logics) Unfinished manuscript, John von Neumann Archive, Library of Congress, Washington, D.C. reviewed by A. H. Taub in: *Collected Works Vol. 1 Logic, Theory of Sets and Quantum Mechanics*, A. H. Taub, ed., Pergamon Press, pp. 195-197.

Weyl, H., 1944, David Hilbert and his mathematical work, *Bulletin of The American Mathematical Society*: 612-654.

Wightman, A. S., 1976, Hilbert's sixth problem, *Proceedings of Symposia in Pure Mathematics*, AMS **28**: 147-240.

CLASSICAL INDIAN MATHEMATICAL THOUGHT IN THE CONTEXT OF THE THEORIES OF MATTER AND MIND

NAVJYOTI SINGH
Indian Institute of Advanced Studies, Rashtrapati Niwas, Shimla, India

Abstract: It is said that – if mathematics was considered as queen of sciences in Greece then linguistics was considered as queen of sciences in India. There is a noteworthy absence of mathematical physics in the Indian mathematical traditions. On the other hand mathematical thought was employed for understanding working of mind by different Indian philosophical schools. We will explore reason for this strangeness by focusing on the nexus between the ideas of causation and mathematics in the classical Indian intellectual context. The relation between causation and mathematics is clarified through the causal analysis of numeric cognition. It is shown that the insights thus gained can be generalised to causally account for any cognition.

Key words: causation; mathematics; extension; obstructability; causal asymmetry; expectancy; adequacy; cognition.

1. ENIGMATIC NEXUS BETWEEN CAUSATION AND MATHEMATICS

The two popular doctrines that continue to impinge upon concerns regarding the relation between physics and mathematics are:

1. *Causal closure of physical domain*: Kim has articulated 'causal closure' disposition as (Kim, 1993) – "... the assumption that if we trace the causal ancestry of a physical event, we need never go outside the physical domain."(p. 280)

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 55–72.

2. *Causal inertness of mathematical domain*: The 'inertness' disposition can be characterized in the words of Balaguer (Balaguer, 1998) as – "… the belief that there is something real and objective that exists outside of space-time and that our mathematical theories characterize."(p. 8)

These two doctrines are seemingly founded on independent rationale and practice. Mathematics is seen as causally inert and physics as causally self-contained. Together they radically partition domains of physics and mathematics rather neatly.

In spite of the division between causal order and logical order that seemingly informs physics and mathematics respectively, we have witnessed mutually enriching and close collaboration between the two disciplines for the last four centuries. The mysterious accord between the two is strikingly illustrated by Galileo's purely logical proof of an entirely empirical issue, namely, the proof of "equal rate of descent of material bodies."[16] He arrived at a causally significant feature of gravitational reality on the basis of purely a-causal and logical argument. Similarly, purely for the sake of mathematical brevity, Dirac paved way for the reality of positron and anti-particles. These examples show that the relation between causality and mathematics is not as innocent as hinted in the powerful twin-doctrines stated above.

At a fundamental level these twin-doctrines were reinforced by convergence and congruence of two independent philosophical distinctions, namely, between

1. Causality and logicality: in the Indian philosophical traditions[17] similar distinction is drawn between *pramāṇa* (veridical causation) and *tarka* (eliminative reason), and

[16] We quote from Galileo to bring home the import of his proof (Galilei, 1632) – "But, even without further experiment, it is possible to prove clearly, by means of a short and conclusive argument, that the heavier body does not move more rapidly than a lighter body."(p.62) The proof, given by Galileo, proceeds by taking two bodies, say, *A* which is heavier then *B*. If they are made to fall from the same height, either heavier will fall earlier or lighter. In case heavier falls faster, tie up with mass-less string the two masses as *A*•*B*. When *A*•*B* falls from the same height, *A* will pull *B* down whereas *B* will pull *A* up and as a result *A*•*B* will fall in between the time taken for *A* and *B* to fall the same height. But since *A*•*B* is heavier then *A*, it will fall faster then *A*. Thus there is a logical contradiction. Same contradiction will result in the case when lighter body falls earlier. To avoid contradictions all bodies will have to fall with equal rate of descent. QED.

[17] *Pramāṇa* is derived from *pramā* (true cognition) + *karaṇa* (causing of) meaning 'causation of true knowing'. This is distinguished from *tarka*, suppositional reason, that does not directly yield 'true knowledge' but only helps focus causal apparatus of 'phenomenal knowledge' by eliminating logically contradictory possibilities. *Tarka* is a-causal whereas *pramāṇa* is not. *Tarka* radically leaves open possibility of new 'phenomenal truth' about

2. Matter and mind: in the Indian philosophical traditions[18] similar distinction is drawn between *jaḍa* (stuff) and *cit* (consciousness).

Though these philosophical distinctions are very old, both in the Greco-European and the Indian analytic traditions[19], *their convergence is new* and has been brought into force only since European Renaissance. Basic impetus for this convergence came from Descartes' radical bifurcation[20] of spatio-

---

its subject matter. For example, Galileo's argument in *footnote 1* is an instance of a particular *tarka*, *reductio ad absurdum*. It, however, presumes identity of the nature of vertical and horizontal motion (that is, unity of gravitational and inertial mass) to derive contradictions, which is questionable as was later done by Einstein. Thus, General Theory of Relativity, giving new causal knowledge about gravity, can be articulated in spite of this *reductio* argument. *Pramāṇa* stands for causal aspect of knowing and not inert causality of matter as such. More accurate translation of *pramāṇa* is 'causation of veridical cognition'. The distinction between *tarka* (eliminative reason) and *pramāṇa* (veridical causation) is firmly upheld by various Indian philosophical schools except by the Jaina tradition, which regards *tarka* as an independent *pramāṇa*. See (Singh, 1997) for fuller discussions on this Indian distinction, which is homologous to the distinction between causal and logical order but can be more accurately rendered as a distinction between 'causation of veridical cognition' and 'ratiocination'. In the Indian traditions reflection on causal aspect of matter is subsumed under causal aspect of veridical cognition.

[18] Ancient Vedic doctrine of *ātman* (self) created a deep and lasting schism from *bhūta* (matter). Separation of the two has been internalized in a most general and widely accepted distinction between *jaḍa/cit* (stuff/consciousness) in the Indian philosophical traditions. The specific characterization of the two, however, differs from one philosophical school to another. Usually 'phenomenal mind' (*manas*) is understood as *jaḍa* (stuff) along with *bhūta* (matter) and both are radically distinguished from *cit* (consciousness). Even *anātmavādin* (no-self-theorist) Buddhist accept this distinction and most radical among them identify transcendental *cit* (consciousness) with *śunya* (emptiness).

[19] Various articulations of these distinctions are well known in the Greco-European traditions as well as in the Indian traditions. A recent comparative survey of Indian and Greek philosophies, (McEvilley, 2002), shows antiquity of the distinction between matter and mind in Indian and Greek philosophies. However, distinction between causal and ordinal aspect of reality in the Indian tradition is not well known. It is Praśastapāda (530, p. 238, 247 & 272-313) who first proposed that numbers do not play any causal role though they are effects that are caused and that they involve event of 'expectancy cognition' (*apekṣā buddhi*) as a conditional cause for their production and that numbers play role in revealing/making order in reality. The distinction between causal role and ordinal role has no obvious relation with various characterizations of the distinction between matter and mind in the Indian analytic traditions including the one proposed by Praśastapāda.

[20] Concept of *res extensae* was propounded by Rene Descartes to create lasting distinction between the material world and the mental world. "Matter has the essential attribute of extension, and all genuine properties of matter, must be (quantitative) modes of extension. These modes include duration, which necessarily is contained in our conception of the existing material things, since to conceive of it as existing is to conceive of it as continuing

temporally extended matter and un-extended mind (Descartes, 1641). Causality was thought to be necessarily and only associated with spatio-temporally extended stuff and not with non-extensional reality. Numbers, however, are not given to us the way spatio-temporal matter is given to us. Numbers in their being are mind-entangled and are non-extensional (in the sense of non-spatio-temporal-extension). Likewise, in logic, entities like negation etc. are non-extensional. *On the basis of the property of extension, the matter-mind dichotomy becomes intimately congruent with causality-logicality dichotomy.* It is only this congruence that remoulded ancient philosophical distinctions into twin-doctrines pointed above.

2. CAUSALITY BEYOND MIND-MATTER INCOMMENSURABILITY

However, there is a serious error in Descartes' extension doctrine. Mind is extended as much as matter is, though *extension of mental entities is purely temporal* whereas *extension of material entities is spatio-temporal*. Mental entities are temporally extended[21] at least in two senses –

(i) mental events ceaselessly consecute, and
(ii) mental event can be composite and thus can be inclusive of the content from temporally different mental events.

Further, relative temporal indexing of mental events is possible – as in reports such as "first I thought, then I felt bad and then I remembered etc." Though, unlike material events, mental events are not spatio-temporally extended, as there is no place for them to dwell (perdure) synchronically and to be commeasured as coeval real events. *Nonetheless any extended entity is obstructable or transversable, hence manipulable and evidentially theorizable*.[22] Mental as well as physical entities are prone to obstruction, hindrance, cessation, sublation and termination and thus ought to be subject

to exist." (Descartes, 1644, I, 55 & 57). According to Descartes, duration of mental event/entity is *inconceivable*. Thus in contrast, the realm of mind, *res cogito,* is non-material without any attributes of extension.

21 In contrast to the Cartesian outlook, modern phenomenological traditions, Brentano and Husserl *et al*, in Europe have always accepted pure temporal extension of mind and consciousness. In the Indian philosophical traditions experienced (*anubhuta*) mental entities are universally regarded as extended in time.

22 Detailed arguments, drawing from Indian analytic traditions, on mind being obstructable (*bādhita*) as much as matter is obstructable has been given in Singh (2003, ch.4) and a generalized theory of obstruction is proposed using which causal underpinnings of the laws of motion or consecution for first person experience have been suggested in the book.

to causation. Events of both classes are at least subject to one or other caused obstruction in their persistence. More generally, they are amenable to being caused into birth, existence and termination. By virtue of being caused, mental as well as physical entities/events are manipulable.

For Descartes, manipulability rests on durability (an extension "which necessarily is contained in our conception of the existing material things, since to conceive of it as existing is to conceive of it as continuing to exist", see footnote 5). Unlike physical events mental events are not durable; they are momentary (they have intrinsically indeterminate duration). It needs to be realised that manipulability results not because of durability but because of the terminals of durable. Eternally durable entity is neither caused into birth nor caused into termination. Non-eternal entities are caused at the junctures of beginning and end terminals. More fundamental are the terminals of duration and not duration itself. Duration can be indeterminate in spite of having determinate terminals. Extension is simply not a key that lets in causation; instead obstructability is a window to causation. *Descartes' error is in mistaking powers of the terminals of extension for powers of the extension*. Mental as well as physical events/entities are subject to obstruction and thus to causation. Therefore, the distinction between matter and mind cannot exactly be congruent with the distinction between causality and logicality. Mental events are not causally inert, nor are physical events logically inassessable. Thus, there is a serious breakdown of Renaissance congruence that had originally enshrined the twin-doctrines. This collapse throws up a challenging arena for fresh reflection on the relation between causality and mathematics. It can be safely claimed that *the nexus between causation and mathematics has to be re-addressed in the context that clearly rises above matter-mind incommensurability thesis*, which is the basic force behind the twin-doctrines. Traditional Indian analytic and conceptual milieu precisely provides such a context. Causal aspect of first person experience and in particular causal aspect of cognition has been a subject of sustained inquiry and abiding interest in the Indian analytic traditions.

In the Indian analytic traditions the *hard* impenetrable mental-material dichotomy is never proposed though a firm distinction between mental and material is maintained. Mental and material is resolutely distinguished on the epistemic ground that mental is privately accessible to one cognizer whereas material is publicly accessible to all cognizers. At least Cartesian extensional dichotomy is not upheld. Instead, hard-dichotomy between obstructable (*bādhita*) and unobstructable (*abādhita*) entities[23] is accepted by numerous

[23] The distinction between obstructable (*bādhita*) and unobstructable (*abādhita*) entities is an analytic distinction. Obstructability or 'being transversable' is a *property of being subject*

contending and persistently quarrelling Indian philosophical traditions. And for all of them, almost without exception, *phenomenal mind (manas) as well as matter firmly belongs to the realm of obstructable entities* (*jaḍa*). For, it is in the nature of mental and material entities that they are subject to temporal consecution and spatio-temporal change respectively. Such change is not possible without causation. Both types of entities are non-perennial, since their perdurance is contained, checked or obstructed by the rest of the world. They are subject to causation and change irrespective of whether they are publicly accessible (as are material entities) or privately accessible (as are mental entities). The very conception of the realm of obstructable entities simply breaks down mind-matter incommensurability thesis.

In contrast to the obstructable realm is conceptualized the realm of unobstructable.[24] There is no necessity that unobstructable entities/entity cannot play causal role, however, it is necessary that they are not effects of any cause. Obstructable entities, in contrast, can be cause as well as effect.

The Indian framework of obstructability provides a natural platform for the study of the nexus between causality and mathematical objects. In particular, it is in the perennial-pluralist outlooks, like that of Vaiśeṣika and Jaina traditions, that causality-mathematics nexus is explicitly worked out.

---

to containment in the state of perdurance. It embodies most elemental conceptualization of 'change' that is common to mental and material entities. Ceasing to perdure involves causality; such an entity requires external causal condition for it to be non-perennial. It is not perdurance as such that involves causality, as Cartesians would want to believe. For, perennially perduring entity is not caused, only an entity that is obstructable is caused.

[24] Usually *Brahman* (often translated as 'pure consciousness') and even *Māhākāla* (grand time) are popularly conceived as unobstructable entities. Śankara's *Advaita* (passive non-dualism), Śaiva *Advaita* (active non-dualism) Bauddha *Śunyavāda* (emptiness-monism) are the three popular philosophical schools that have cultivated monist idea of unobstructable entity. Vaiśeṣika, Mīmāṃsā and Jaina philosophical traditions take pluralist position on unobstructable entities and propose comprehensive system of invariant categories. Even plural *padārtha*-s (scheme of ontological categories) can be legitimately conceived as unobstructable entities that perennially perdure and remain invariant through changes. Praśastapāda is an important Vaiśeṣika thinker who explicitly theorized on the problem of the nexus between causality and mathematical objects (Praśastapāda, 530). In such a reckoning, space and time are unobstructable real entities among other such entities, whereas all events, mental or physical, are obstructable. Analytic issues such as the unity or plurality of the realm of unobstructable entities or issues such as causal/logical relation between obstructable and unobstructable entities are ceaselessly debated by traditional Indian philosophers and theorists.

3. CAUSATION OF MATHEMATICAL ENTITY IS MIND ENTANGLED

In modern scholarship Benacerraf[25] (1983, pp. 403-420) has argued, "If mathematical facts are causally inert, we cannot know them (or entertain belief about them)." Maddy has suggested the realist causal underpinnings of the perceptibility of sets of physical objects (Maddy, 1990). Koons has proposed a causal theory of modal knowledge, including logical and mathematical knowledge (Koons, 1999). Mathematical knowledge itself is causally arrived at simply because mental activity involving mathematical entities is subject to causation. Phenomena of causally governed mental activity are assertably true even from the first person perspective apart from being true from the third person perspective. *Mathematics is thus mind-entangled in its causal bearings*. But the question is in which precise way numbers and other mathematical entities can be causally accounted for. More compelling is the question whether mathematical truth itself is causally anchored. And what endows mathematics with the power to play a role in causality-centric physical theories.

Praśastapāda (530, p. 239) had proposed that numbers, numeral-distinctions and graphic-aggregations are fundamentally mental (*buddhyapekṣa*) in nature.[26] Their *being* involves mental event and mental properties. They come into being because of mental activity. Śridhara (991, p. 239), in his commentary on the passage, gives a following supportive syllogism – "The number 'two' is produced by cognition because 'two' is cognized by only one cognizer; All that are cognised by only one cognizer are produced by cognition like pain etc.; The number 'two' is cognised by only one cognizer therefore 'two' is produced by cognition." This argument, however, does not diminish objectivity of 'two' that could be independently cognised by each of many cognizers. The argument can be generalized as – "mathematical entity is produced by cognition, because while being an effect it is necessarily cognised by one cognizer, like satisfaction." Thus, *causation of mathematical entities is mind-entangled*. Mind is necessarily implicated in the coming-to-be of mathematical entities and interestingly such coming-to-be is not a material cause of anything.

[25] Earlier too Benacerraf had made a plea for the causal aspect of mathematical object (Benacerraf, 1965).

[26] *paratvāparatvadvitvadvipṛthaktvādayo buddhyapekṣā.*

4. CAUSAL ASYMMETRY INVOLVED IN MATHEMATICAL ENTITIES

Interestingly, in the Vaiśeṣika analytic tradition, Praśastapāda (530, p. 247) had claimed that numbers (*dvitvādi saṃkhyā*), numeral-distinctions (*dvi-pṛthaktva*) and graphic-segregations (*paratva-aparatva guṇa*-s) are not cause of anything unlike substances (*dravya*), actions (*karma*) and other qualities (*guṇa*), which play determinate causal role. Such non-causative entities can be safely called pure mathematical objects. This holds for all numbers other then 'one', which, however, plays a role of a cause.[27] At the same time he had maintained that numbers (other then 'one') etc. are caused and thus are effects of determinate causes.[28] This is an interesting thesis of causal asymmetry involving mathematical entities. They are effects but not causes. Unlike unobstructable entities, which cannot be effects, numbers etc. are effects. Among obstructable entities, numbers etc. are distinct from the

[27] Praśastapāda (530, p. 243) states that number 'one' inherent in each of the constituent atoms, is a cause of 'one' in a whole made out of them and is also a cause of 'two' etc. in loci of the parts of whole. Such a causal theory is a result of sophisticated Vaiśeṣika doctrines of 'whole residing in each of its parts' and 'whole being constituted by parts and yet being different entity from sum of parts'. However, caused numbers 'two' onwards are not further cause of any being. Number 'one' creates effect in its own locus (like 'two' etc.) as well as in other locus (like 'one' in a whole). This is said to hold for the causal role of one-distinct-ness (*eka-pṛthaktva*) as well. Praśastapāda (530, p. 240) says that unlike several real qualities, number 'one' and 'one-distinctness' produce effects of the same kind, i.e., numbers and distinctness respectively. Another property that goes with 'one' and 'one-distinctness', according to Praśastapāda (530, p. 249), is that they remain till their loci exist whereas other numbers etc. can disappear while their loci survive. Because of these complications with 'one' Jaina thinkers even upheld that smallest number is 'two' and not 'one'.

[28] Apart from causing numbers in physical-whole there is one other way that number 'one' plays a causal role. This deals with the construction of mental-whole. Mental-whole, i.e., cognition, is different from physical-whole *only* in a way whole relates to its parts. A single perceptual cognition has many parts or aspects that ultimately constitute it, but such constituted cognition does not inhere in those parts whereas physical-wholes necessarily inhere in their parts. When I visually perceive 'monitor, table etc.' as mental-whole, even if my head moves towards left to visually perceive 'bookshelf, papers etc.' as another mental-whole, the monitor etc. are not left poorer because of lack of visual cognitive-whole that they had constituted and of which they no longer are parts. 'One', subsisting in a real locus, can cause perceptual cognition 'this is one'. Two such cognitions in consecution lead to mental event involving an entity 'two-ness', which in turn causes 'two' in the constituents of perception. Such an entity 'two-ness' is available to self (cognizer) as its surrogate property (*dharma guṇa* of *atman*) and the process is quite akin to Platonic reminiscence of form by soul. Subsequently, perceptual cognition 'these are two' is caused.

rest by not being a cause but being only effects. Thus there can be a division of entities into three classes on the basis of their role as a cause and effect –

(i) causes but never effects: unobstructable entities;
(ii) effects but never causes: obstructable mathematical entities, and;
(iii) causes as well as effects: other obstructable entities.

Praśastapāda (530, p. 231) had also noted that the unique causal asymmetry of mathematical objects (i.e., being effect and not cause) endows them with the ordinal power over everything including them. He had included them in the category of general qualities (*sāmānya guṇa*), which are found in all substances without exception. They do not causally soil the reality but are nonetheless created by the reality and in turn disclose ordinal features of reality. Product-cum-causally-impotent nature endows numbers etc with supervening and divulging role.[29]

Again this does not hold for 'one' whose being is not dependent on mental processes and thus can be a cause as well as an effect unlike other mathematical entities. 'One' can produce as effects numbers in two types of wholes – (i) physical-whole (objects), which inheres in each of its parts, and; (ii) mental-whole (cognition), which does not inhere in its parts.[30]

Causal behaviour of 'one' as a cause and as an effect is implicated in the process of the construction of physical-whole from parts. Formation of a unitary physical-whole, in which comes to in-exist 'one', is explained by Praśastapāda with a help of a general Vaiśeṣika thesis (*ārambhavāda*) that "being of whole though constructed by parts is different from the sum of parts." Though its parts causally construct a physical-whole, it becomes a whole only if it comes to in-exist in each of its parts. Inhering of a real physical-whole in each of its parts endows it with robust compactness and unity. Unity of a physical whole, as a quality 'one' in-existing in a whole, is produced by (i) 'one' inhering in each constituent parts, and (ii) 'whole' also

[29] In contrast, popular modern doctrine – 'Platonic realism of numbers', see (Balaguer, 1998) – takes numbers as perennial beings, which can be causes but not effects. It is thought that matter is condemned to imitate such perfect beings. Such a meek imitation by uncouth matter supposedly accounts for the potency of numbers. Reminiscence of their being leads to numeric cognition as well as to true verifiable knowledge about material form.

[30] In Western philosophy, Brentano (1917, p. 195), using Aristotelian terminology, makes a similar distinction between two types of accidents – (i) An accident that is *last-of-the-part-full* is *inherent accidents* (*eigenschaften*) that inheres in its subject (in all its parts) and is a compact whole which displays unitary coherence of parts when acted upon and does not require activity to sustain itself; (ii) Accidents that are *mere-part-full* are called by Brentano *passive affections* (*erleidungen*) which require activity to remain in their subjects. These exactly correspond to physical whole and mental whole in our terminology.

inhering in each of them.[31] Through such intricate mereology Praśastapāda explains how a physical-whole moves when only one part of it is pushed or how entire physical-whole is perceptually cognised while only one portion of it is visible to the eye or is touched. Unity of whole, embodied in a produced quality 'one', clearly has a causal role there.

5. "EXPECTANCY COGNITION" IN THE PRODUCTION OF NUMBERS

Production of numbers other than 'one' in a mental-whole and in a physical-whole is an altogether different process. The complexity in understanding these processes results because mental-wholes do not in-exist in their parts. Locus of mental-wholes is self and not what is cognized. Further, numbers other then 'one', even in physical-wholes, for their existence presume occurrence of a mental-whole. Physical-wholes occur as parts of a mental-whole. Causal production of numeric cognitions by 'one' in-existing in such parts of a mental-whole is analytically a challenging issue. The simplest case is the production of 'two' in physical-wholes (as in "'two' things") and its perceptual cognition "two things". According to Praśastapāda this process involves subsequent atomic cognitions "this is one" and "that is one". 'Two' as an entity is produced and destroyed in the process of the formation of a final mental-whole (that is, perceptual cognition) "two things". Several cognitive events take place in causal order to accomplish this process. Working out the causal details of numeric cognition, Praśastapāda (530, p. 272) had proposed that it is invariably an episodic event of 'expectancy-cognition' (*apekṣā buddhi*) that brings them (numbers other then 'one' etc.) into being.[32] He had worked out a detailed causal process involved therein (pp. 272-313).[33] This causal process of

[31] According to Praśastapāda there are certain qualities that produce effect in locus other than their own and 'one' is such a quality. 'One' inhering in part produces 'one' inhering in a whole. When a physical-whole comes into being it comes to inhere in each of its parts. In any part, apart from inherence of substantive whole, also inheres a quality 'one'. These in-existing 'one's of all parts simultaneously, through a binding act, produce a quality 'one' in a physical-whole. This is because in each part it is the same whole that comes to reside. Foundations of this mereology have been formally articulated in Singh (2003, chapter 5).

[32] Commentator Śridhara (991, p. 272) had clarified that in the production of 'two', two substances are the material cause, two of the numbers 'one' are the non-material cause and 'expectancy cognition' is the efficient cause.

[33] The entire process of the production and the destruction of 'two' in perceptual cognition "two objects" involve six serial moments (*kṣaṇa*-s) and four cognitive episodes (*jñāna*)

mathematical cognitions has been a subject of discussions, criticism and suitable amendments from different perspectives by several thinkers since then.[34]

Since production of 'two' etc. is a result of the conditional existence of 'expectancy cognition', numbers in their being are non-persistent as 'expectancy cognition' is non-persistent. Ephemerality of cognition permeates down to ephemerality of its effects. So far as numbers (other than 'one') are qualities of substances and numbers are caused effect, *numbers have 'contingent being' dependent on 'expectancy cognition'*. Number, however, has many senses apart from the sense of the quality of substances. It is these other senses of number that can make it possible for even robots to count.

6. THREE-TIER ONTOLOGY OF NUMBER

Interesting to note in the proposed causal process is that Praśastapāda delineates and employs three different senses of a number. For instance, nominal number '2' has following three ontological senses[35]

(i) 'two-ness' as an abstract entity (*dharma*);
(ii) 'two-ness' as quality (*guṇa*), and
(iii) 'two-ness-ness' as a natural kind (*jāti*).

He had stitched together these three senses of number into an episodically unified causal process of counting, which eventually produces perceptual cognition "two physical-wholes". Nominal number '2' as a quality is technically called 'two-ness' (meaning ordinal 2) because it independently in-exists in two different loci of two things at once. All possible occurrences of such 'two-ness' are however instances of a universal or natural kind 'two-ness-ness' (meaning cardinal '2'). In distinction from these two types of real entities, namely quality and universal, 'two-ness' as

---

with each cognitive episode of three moments. In the process quality 'two' comes to in-exist in real substantive objects before being perceives in cognition.

[34] For instance, according to the analysis of Śankaramiśra (1430, pp. 219-223) the entire process of creation and destruction of 'two' takes 16 moments.

[35] Praśastapāda's cognitive process of counting is: (i) Perceptual cognitions of "'one-ness' inherent in 'one' inhering in a physical-whole" and of "'one-ness' in 'one' inhering in another physical-whole"; (ii) Expectancy cognition of "abstract 'two-ness' in each 'one'"; (iii) Birth of quality 'two-ness' in each physical-whole; (iv) Perceptual cognition of universal 'two-ness-ness', and (v) Perceptual cognition of 'two-ness' quality, and (vi) Perceptual cognition "two physical-wholes". The existence of 'two-ness' quality in physical-wholes lasts till 'expectancy cognition' lasts.

an abstract entity is necessary to account for the causation of the content of 'expectancy cognition', which indeed is a crucial part of the causal process. Since veridical cognitions "this is one" and "that is one" cannot by themselves provide an entity '2', an abstract 'two-ness' has to be posited. This abstract 'two-ness' is not a cause but a content of 'expectancy cognition'. From where does an abstract entity 'two-ness' arrive in the first place?

Ontologically, abstract 'two-ness' is always available to cognizer as a 'non-experienced' quality of self (*adṛṣṭa guṇa* of *ātman*). It is called abstract (*dharma mātra*) because though it is located in self it also can surrogate to be a property of an entity (*dharmi* or its adopted locus) that is different from the self. Abstract entities can be defined as entities being capable of *parālambana* or 'embracing the locus other than its own'. Vaiśeṣika term for them is *aupādhika dharma* (imposable properties). Abstract number merely dangles in the self, fated to be embodied in substances through 'expectancy cognition' while counting. Various qualities of self usually are expected to characterize self, but abstract entities, though being a special quality of self, come by its nature to characterize entities other than self.[36] Besides, fund of abstract entities dwell in self without being experienced. Memories, for example, are abstract entities that are non-experiential quality of self (*adṛṣṭa guṇa* of *ātman*). They however participate in experiential 'memory cognition'. Abstract 'two-ness' is similarly recollected from non-experiential fund of self in an event of 'expectancy cognition'. Experienced qualities are temporally extended, are ephemeral and are momentary (like cognition, desire, aversion, effort, contentment etc.) but non-experienced qualities (memory, morals, classes etc.) are synchronic depositories. They do play a-causal role in constructing experience like that of "two wholes".[37] The depository of abstract entities is, however, caused by the traces of experience.

[36] Among 'non-experienced' qualities of self are entities that characterize self itself. These are moral dispositional qualities like merit/demerit (*dharma*/*adharma*). Like other abstract entities, which embrace other locus as well, moral dispositional qualities necessarily implicate imposition on the entire expanse of the plurality of selfs and not just a particular self. Nature of *parālambana* in moral disposition is an interesting issue that is intimately related to the universal role of abstract entities.

[37] Such a role of abstract entities in cognition is quite analogous to relation between *tarka* (eliminative reason) and *pramāṇa* (veridical causation) noted in *footnote 2*. *Tarka* is like memory that does not directly yield 'true knowledge' but only helps focus causal apparatus of 'phenomenal knowledge' by endowing it with parsimony. Parsimonious function in mind is non-experienced (*adṛṣṭa*) but underlies all experience (*anubhava*). Production of abstract entities as quality of self is an independent and interesting causal question. In a general way it can be said that any event of structured cognition (*savikalpaka jñāna*) goes

The idea of an abstract number is important as it plays logical role in the causal construction of ordinal and cardinal number. It plays role in the content of 'expectancy cognition', which in turn causes in other real objects a momentary existence of quality (ordinal number) and natural-kind (cardinal number) for the duration of its existence. These momentarily existing ordinal and cardinal in real locus give rise to perceptual cognition with numerical content. This numeric cognition in turn causes formation of abstract depository as its trace. Thus, *the logical and the causal are integrated in a loop-like manner* in Praśastapāda's causal process of the cognition of number. The key step in the process is occurrence of abstract or imposed property (*aupādhika dharma*) in 'expectancy cognition'.

Nexus between causality and mathematical entities is one of the most intricate theories proposed by Praśastapāda involving –

(i) causal asymmetry of mathematical entities;
(ii) mind-entangled causation of mathematical entities;
(iii) necessary occurrence of 'expectancy cognition' in a production of mathematical entity;
(iv) ephemeral nature of real mathematical entities;
(v) existence of abstract mathematical entities;
(vi) three-tier ontology of a mathematical entity, and;
(vii) loop-like integration of logical and causal in the production and the destruction of mathematical entities.

In the proposed causal process, the realm of abstract entities and the realm of real entities participate in an ordered and causally exact manner.

7. ADEQUACY CRITERION AND THE PROBLEM OF THE UNITY OF MENTAL WHOLE

It was felt that in this account of numbers there is one issue that has been left unaccounted for. The issue is regarding production of the unity of cognitive-whole, in particular production of the unity of 'expectancy cognition'. Production of unity in physical-whole can be understood and such a unity leading to cognitive generation of numbers in things can be

---

in the production of abstract entities. *Vijñānavādi* Buddhists have paid detailed attention to this causal process. In modern research Zalta (1983) attempted a-causal axiomatic theory of abstract entities and Tennant (1997) gives *a priori* arguments for the necessity of abstract entities. Contemporary disputes over existence of abstract entities are reviewed in (Burgess and Gideon, 1997).

understood as well. But how does one understand unity of 'expectancy cognition'?

Raghunātha (1510) gave a deep turn to Praśastapāda's theory by proposing a relation of adequacy (*paryāpti sambandha*) to understand unity of cognitive-whole.[38] Number 'one' does not capture the unity in a hypothetical construction "one expectancy cognition", since 'one' in such cognition is not caused by the content (parts) of cognition. Instead, *paryāpti*[39] (adequacy) relation between abstract entity and real entities captures compact adequacy[40] of 'cognition'. In case of 'expectancy cognition', adequacy relation obtains between an abstract 'two-ness' and real 'two-ness' that inheres in each thing that is paired. The real 'two-ness' that inheres in each member of pairs is a 'class of two members', that is, each member inheres that class to which it has come to belong. Such a class is thus compact because mental act, where abstract 'two-ness' occurs, makes real members belong to that class. Abstract 'two-ness' is instead 'class of all classes of two members' which is related by adequacy relation to the pairs and not to the members of pairs.

In modern period Frege similarly *defined number exclusively as an abstract entity* since for him nominal number '2' is 'class of all classes of two members'.[41] But Raghunātha's adequacy relation is a cognitive structure relating abstract 'two-ness' with pair of two entities in each of which inheres real 'two-ness' as quality. It is adequacy relation that makes loci of abstract 'two-ness', 'three-ness', 'four-ness' etc. as mutually exclusive even if same real entities participates in them. For Raghunātha *number is a structure governed by adequacy criterion that relates abstract number with real entities. Mathematical knowledge is an expression of adequacy conditions associated with numerical cognitions*.

The theory of adequacy relation is later generalized by Gadādhara (1660) to cover all cognitions and not just restricted to application in understanding structure of numeric cognitions. In any cognition, meaning of participating real entities is limited (*avacchinna*) by other participating real entities.

[38] While defining an analytic device of limitor-ness (*avacchedakatva*) Raghunātha (1530, pp. 42-43) introduced an idea of adequacy relation. "The limitor must be understood as occurring in its locus, i.e., as occurring completely in it by the adequacy (*paryāpti*) relation." Earlier even Bhāsarvajña (950) had flouted Praśastapāda's theory on the ground that even "'two-ness' is 'one'."

[39] *Paryāpti* means 'completion', 'wholeness' (*paryavasānam, sākalyam*).

[40] In temperament, the idea of 'adequacy relation' is somewhat akin to Tarski's theory of truth (Tarski, 1933).

[41] This was pointed by Ingalls (1951) and later analysed by Shah (1982), Roy (1985) and Matilal (1985).

Usually meaning of a real entity is determined by the scope of a real universal that characterizes it and inheres in it. But in cognition this scope is curtailed by other participating entities. Each such curtailment is called limiter (*avacchedaka*). It is in this mutual-curtailment of participating entities that unity or 'unified binding' of that cognition has to be founded. Cognitive-whole is unified and complete iff 'extent of limiters' (*avacchedakatā*) of participating entities (their meaning may not extend to whole extent of their corresponding universals) gets related by adequacy relation with abstract limiter-ness-ness (*avacchedakatātva*). It is such an adequacy that makes cognition a well-bounded whole. Gadādhara (1660) employs this generalization for explaining how perceptual cognition like "smoke on the mountain" through adequacy relation causes inferential cognition "fire on the mountain." *Logical knowledge is an expression of adequacy conditions associated with any cognition*.

Though arguments given above deal with numbers but they can be appropriately recasted for numeral-distinctions and graphic-aggregations which also invariably involve 'expectancy cognition' in their causal account. 'Adequacy relation' would relate corresponding abstract entities with 'co-ordinated' real entities to understand causal underpinnings as well as to understand binding composure of cognition. Mathematics is not about abstract entities alone but is about relation of abstract entities with real entities. Further, though the example taken for analysis above is regarding causal account of simple cognition "two things", the analysis holds for all numbers with suitable adoption. *Prime object of mathematical investigations is adequacy conditions that implicate mathematical operations, constructions and application to variety of real situations*.

Adequacy relations between abstract and real entities provide space or opportunity where mathematical and logical thought operates parsimoniously. Solution of a typical mathematical problem (say by a student) is cognitive movement form the situation of inadequacy to the establishment of adequacy relation demanded by the problem. Exact formulation of inadequacy (for instance with the idea of variable in many typical algebraic school problems) and its solution typically takes place over several cognitive episodes (even interspaced with faltered cognitive episodes). Once adequacy relation is established, resolution of such problems can take place even in a single cognitive episode.[42] It is real

[42] Indian mathematician Srinivasa Ramanujan (1887-1920) with no university education would just list results on elliptic functions, continued fractions and infinite series without being able to reproduce procedure of arriving at the results. When Hardy asked him as to how he arrived at results, his explanation was no more then that the Goddess reveals it to him. Hardy remarked (Hardy, 1940) – "The limitations of his knowledge were as startling

situation of inadequacy (as is witnessed in familiar mathematical, physical and social problems) that drives mathematical and logical thinking. Invention of new adequacy relation punctuates history of ideas and is also a task wide open at the frontiers of ideas.

8. CONCLUSION: NEXUS BETWEEN CAUSATION AND MATHEMATICS

Foundation of mathematics outside the spell of mind-matter incommensurability thesis involves causal account of mathematical cognition and mathematical entities. Riding on the shoulders of Vaiśeṣika tradition we have conceptually cleared ground for that. Mathematical entities have three-tier ontology apart from their nominal being. These are woven together in adequacy relation between abstract and real that obtains in any mathematical cognition. Mathematical theory is an organism of such adequacy conditions imbedded into a systemic organism, which in turn makes causation of advanced mathematical cognition possible. Rather man's quest for non-experienced systematisations of adequacy relations makes possible causation of any wholesome cognition. Abstract depositories adequately copulate with reality to yield normal cognitions, which in turn causes reconfiguration of abstract depository. Mathematics is discovered and yet invented because it is inalienably soiled with causality.

REFERENCES

Balaguer, M., 1998, *Platonism and Anti-Platonism in Mathematics*, Oxford University Press, Oxford.

Benacerraf, P., 1965, What numbers could not be, *Philosophical Review* **74**: 47-73. Reprinted in: 1983, *Philosophy of Mathematics*, P. Benecerraf and H. Putnam, eds., Cambridge University Press, Cambridge, pp. 272-94.

Benacerraf, P., 1983, Mathematical truth, in: 1983, *Philosophy of Mathematics*, P. Benecerraf and H. Putnam, eds., Cambridge University Press, Cambridge, pp. 403-420.

Bhāsarvajña, 950/1968, *Nyāyabhūṣaṇa* and auto-commentary *Nyāyasāra*, S. Yogendrananda, ed., Chaukhamba Sanskrit Pratishthan, Delhi.

Brentano, F., 1917/1981, *The Theory of Categories,* translated by Roderick. M. Chisholm and Norbert Guterman, Martinus Nijhoff, The Hague, Boston, London.

as its profundity." His access to 'adequacy relation' was startling. In the Indian tradition an interesting concept of 'one-sentence-ness' (*ekavākyatā*) as a 'systemic organism of adequacy relations' is introduced to explain sudden access to unfamiliar 'adequacy relation' among thinking men.

Burgess, J., and Gideon, R., 1997, *A Subject with No Object*, Oxford University Press, Oxford.

Descartes, R., 1641, *Meditations on First Philosophy*, in: 1997, *Descartes: Key Philosophical Writings*, translated by S. Haldane and G.R.T. Ross, E. Chavez-Arvizo, ed., Wordsworth Classics, Hertfordshire, pp. 123-260.

Descartes, R., 1644, *Principles of Philosophy*, in: 1997, *Descartes: Key Philosophical Writings*, translated by S. Haldane and G.R.T. Ross, E. Chavez-Arvizo, ed., Wordsworth Classics, Hertfordshire, pp. 261-337.

Gadādhara, 1660/1979, *Pañcalakṣaṇa* and *Sinhavyāghralakṣaṇa*, edited, with Raghunātha's *Dīdhiti*, Gadādhara's *Prakāśikā* and editor's *Bālabodhinī*, by N.S.Ramanuja Tatacharya, Rashtriya Sanskrit Vidyalaya, Tirupati, **26**.

Galilei**,** G., 1632/1954, *The Dialogue Concerning Two New Sciences*, translated by Henry Crew and Alfonso de Salvio, Dover publications, New York.

Hardy, G. H., 1940, *Ramanujan*, Cambridge University Press, Cambridge.

Ingalls, D. H. H., 1951, *Materials for the Study of Navya Nyāya Logic*, Harvard Oriental Series **40,** Harvard University Press, Harvard.

Kim, J., 1993, *Supervenience and Mind: Selected Philosophical Essays*, Cambridge University Press, Cambridge.

Koons, R. C., 1999, *Realism Regained: An Exact Theory of Causation, Teleology and the Mind*, Oxford University Press, New York.

Maddy, P., 1990, *Realism in Mathematics*, Clarendon Press, Oxford.

McEvilley, T., 2002, *The Shape of Ancient Thought: Comparative Studies in Greek and Indian Philosophies*, Allworth Press, New York.

Matilal, B. K., 1985, On the theory of number and *Paryāpti* in Navya Nyāya, *Journal of The Asiatic Society of Bengal* **28**: 13-21. Reprinted in: 2002, *Mind, Language and World: The Collected Essays of Bimal Kirshna Matilal*, J. Ganeri, ed., Oxford University Press, New Delhi.

Praśastapāda, 530/1997, *Praśastapādabhāṣya* with Śridhara's (991) *Nyāyakandali*, edited and translated into Hindi by Srī Durgadhara Jha Sharma, Sampurnananda Sanskrita University, Varanasi. Also translated into English by Ganganatha Jha, *The Pandit* (new series), Banaras, **25** 1903 – **37** 1915. Reprinted Chowkhamba, Varanasi, 1982.

Raghunātha, Ś., 1510, *Avacchedakatvanirukti* section of *Dīdhiti* on Gangeśa's *Tattvacintāmaṇī,* edited, with Jagadīśa's *Jāgadīśī* and editor's *Vivṛti* and *Manoramā*, by Vamacarana Bhattacarya, Banaras, 1948.

Roy, W. P., 1985, A note on the Navya Nyāya account of number, *Journal of Indian Philosophy* **15**: 227-234.

Shah, J. L., 1982, Number: from Nyāya to Frege-Russel, *Studia Logica* **41**: 283-291.

Singh, N., 1997, On Tarka: ratiocination and veridication, *Science Philosophy Interface* **2:** 1-20.

Singh, N., 2004, *Mechanics of Phantasm: Antaḥkaraṇa*, (forthcoming).

Śankaramiśra, 1430, *Upaskāra on Kanāda's Vaiśeṣika Sutra*, translated by Nandalal Sinha, Sacred Books of the Hindus, Allahabad, 1911, 1923, Reprinted 1974.

Śridhara, 991/1997, *Nyāyakandali*, commentary on Praśastapāda's (530) *Praśastapādabhāṣya*, edited and translated into Hindi by Srī Durgadhara Jha Sharma, Sampurnananda Sanskrita University, Varanasi.

Tarski, A., 1933, The concept of truth in the languages of the deductive sciences, (in Polish), *Prace Towarzystwa Naukowego Warszawskiego, Wydzial III Nauk Matematyczno-Fizycznych* **34**, Warsaw. English translation, The Concept of Truth in Formalized

Languages, pp. 152-278 in *Logic, Semantics, Metamathematics, papers from 1923 to 1938*, J. Corcoran, ed., Hackett Publishing Company, Indianapolis, 1983.

Tennant, N., 1997, On the necessary existence of numbers, *Nous* **31**.3**:** 307-336.

Zalta**,** E., 1983, *Abstract Objects: An Introduction to Axiomatic Metaphysics*, D. Reidel, Dordrecht.

PART 2

MATHEMATICS AND PHYSICS: REFLECTING ON THEIR INTERACTION

THE ROLE OF MATHEMATICS IN PHYSICAL SCIENCES AND DIRAC'S METHODOLOGICAL REVOLUTION

GIOVANNI BONIOLO[1] and PAOLO BUDINICH[2]
[1]*University of Padova, Padova, Italy;* [2]*The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy*

Abstract: In our paper, avoiding any strong metaphysical commitment on the world, we face the topic of the interplay between mathematics and physics by starting from a semiotic approach. It will be shown that it allows us to insert in a unitary and coherent framework answers to questions such as: Why mathematics is physics? What is the role of mathematics in physics? Why is mathematics effective in physical sciences? In the second part of the paper, and by utilizing what discussed in the first one, we analyse what we call Dirac's methodological revolution, according to which to do good and new physics we must first work on good and promising mathematics. Finally, we exemplify Dirac's methodological revolution by recalling the role of the mathematical theory of simple spinors in constructing new perspectives for theoretical physics.

Key words: mathematics; physics; semiotic; methodological revolution; spinor.

What, however, was not expected by all the scientific workers of the last century was the particular form that the line of advancement of the mathematics would take, namely, it was expected that the mathematics would get more and more complicated, but would rest on a permanent basis of axioms and definitions, while actually the modern physical development have required a mathematics that continually shifts its foundations and get more abstract. (Dirac, 1931)

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 75–96.

1. INTRODUCTION

Certain physicists, when asked why mathematics is so effective as to their theories, reply that they do not know. Others, such as E. Wigner, answer that it is a mystery (Wigner, 1960). Others still, such as F. Dyson, reply that it is not their concern, since they are physicists (Dyson, 1964).

Dyson is perfectly correct in admitting that this problem does not concern physics, but philosophy. Indeed during the history of philosophy many scholars have tried to provide an adequately articulate and argued solution. However, each one may, in one way or another, fall into one of the following lines of thought:

1) the so called Platonic line: mathematics is effective because the world is intrinsically mathematical;
2) G. Galilei's line contained in *Il saggiatore* (1623) and in *Dialogo sopra i due massimi sistemi del mondo* (1632): mathematics in physics is effective because there is a close homogeneity between the physical world and mathematics;
3) G. Berkeley's line contained in *A Treatise Concerning the Principle of Human Knowledge* (1710): mathematics is effective only because it is nothing else but a good tool;
4) I. Kant's line contained in *Kritik der reinen Vernunft* (1781-1787): mathematics is effective because we cognitively constitute the world in a mathematical manner;
5) I. Kant's line contained in *Metaphysische Anfangsgründe der Naturwissenschft* (1786): mathematics is effective in physics because it is only thanks to mathematics that we are able to construct concepts of objects of which we do not have direct experience.

In the subsequent pages, we will endeavor to follow a different path. In particular, we will not assume any metaphysics on the world (as in the Platonic one), or a strongly committing theory of knowledge (such as Galilei's or Kant's), but we will not even demean the importance of the problem by adopting a too naive instrumentalistic approach *à la* Berkeley. Instead, we will start from what Peirce sustained regarding the structure of the physical theories that he analysed in semiotic terms (Peirce, 1895).

Notwithstanding this approach, we will not develop an in-depth semiotic analysis of the physical language. We will limit ourselves to attempting to show how an apparently atypical point of view may describe and highlight interesting aspects of the relation between physics and mathematics. Exactly, by means of this semiotic perspective we will outline plausible replies to the following three questions:

1) *Why mathematics in physics?*
2) *What is the role of mathematics in physics?*

3) *Why is mathematics effective in physical sciences?*

After attempting to tackle the above questions, we will present some remarks on what we call *Dirac's methodological revolution*, according to which to do good and new physics we must first work on good and promising mathematics.

At the end, we will concisely illustrate how the recently emerging mathematical theory of simple spinors while opening new perspectives for theoretical physics, well illustrates the role of Dirac's methodological revolution.

2. WHY MATHEMATICS IN PHYSICS?

First of all, let us ask ourselves whether, as suggested by Quine, it is true that there is no way of doing physics without mathematics (Quine, 1976); whether physics must necessarily contain mathematics, and whether physics has always contained mathematics.

By recalling the existence of Aristotelian physics, which is considered physics without mathematics by excellence, the reply to the last question is immediately negative. Once this is ascertained, it naturally ensues that the first question has an obvious reply. If physics existed without mathematics it means it is possible to have physics without mathematics. Obviously the epistemological and methodological features of that physics are different from those of the contemporary physics. Physics without mathematics is more a philosophy of nature based on intuitive common sense, than a precise and exact science such as those we want now. However, let us not forget that physics without mathematics is not a disorganized pile of data, incapable of any prediction. The Aristotelian physics was a well-organized discipline, albeit not organized according to contemporary canons. In addition, it was also capable of providing predictions. *The absence of mathematics is not at all detrimental to neither predictive power, nor explanatory power, nor organizational power*.

Thus, physics without mathematics is conceivable, or rather, physics without mathematics *was* feasible, because this type of physics is no longer acceptable. The simple reason being that it would not have the empirical and theoretical accuracy which only mathematics provides.

At this point, it is appropriate to recall that Koyré illustrated that, excluding rare instances, the accuracy pertained to Greek astronomy entered physics only as of the XVII century, when began the tendency of using regularly measuring instruments also within the terrestrial domain (Koyré, 1948). Consequently, mathematics, which already existed in ancient

astronomy both in the modeling context and in the predictive context, also began to enter physics enabling it to have greater accuracy.

Notwithstanding Koyré's historical thesis may be criticized, what is relevant from an epistemological view is that it clarifies the fact that when accuracy of a theory of nature is sought, one must mathematize it, since only mathematics allows for the accuracy that, at the most elementary level, is given by the numbers resulting from the measuring instruments.

That the transition *du monde de l'"a-peu-près" a l'univers de la precision* happened in the XVII century, as Koyré suggested, or in a previous time, is the concern of historians of science. What is relevant for us is that this occurred only by mathematizing the representations of nature. The transition *du monde de l' "a-peu-près" a l'univers de la precision* is the transition from a physics without mathematics to a physics with mathematics, that is, to a contemporary physics.

Therefore, at the beginning, mathematics was associated with physics above all as the language that allowed dealing with the numbers linked to the measurements. It is only afterwards that mathematics also becomes an agent for construing well-organized structures from which one can precisely deduce numbers, which will then be compared with the world.

Hence, the least philosophical answer to the question "Why mathematics in Physics?" consists in ascertaining that this is the only way one can obtain a precise physical theory which may be accurately compared with the world.

This is a reply that depends neither on strong metaphysics such as Plato's, nor on strong epistemologies such as Galilei's, or Kant's, but neither does it eliminate the problem instrumentalistically as Berkeley did.

3. WHAT IS THE ROLE OF MATHEMATICS IN PHYSICS?

3.1 The theory of physics as a sign

To propose an *à la* Peirce approach, it is worthwhile briefly recalling that a *sign* is something which connects the *object* with the *interpretant*, that is, both its interpretation (the cognitive aspect of the interpretant) and what has to do with a related action (the pragmatic aspect of the interpretant).

It follows that considering, as we will do, a mathematized physical theory as a *physical-mathematical sign* means regarding it as something connecting the physical world (the object) which we wish to represent, and the

interpretant which, by providing an interpretation of the former, cognitively signifies it and allows us to act in it.

A triadic approach of this kind enables us to deal with many problems in an organized and efficient manner, also because a sign has to be considered as an *icon*, as an *index* and as a *symbol* (fig. l). In other words, three points of view which allow us to consider the *sign as such*, the *sign in relation to the object* and the *sign in relation to the interpretant*, respectively. It is exactly by understanding the significance of the icon, the index and the symbol that it is possible to discern the theoretical value and the mutual dependence of

1) *the physical-mathematical sign as something in itself;*
2) *the relation between the physical-mathematical sign and the world;*

the relation between the physicist and the physical-mathematical sign, as well as the relation, realized by the physical-mathematical sign, between the physicist and the world.

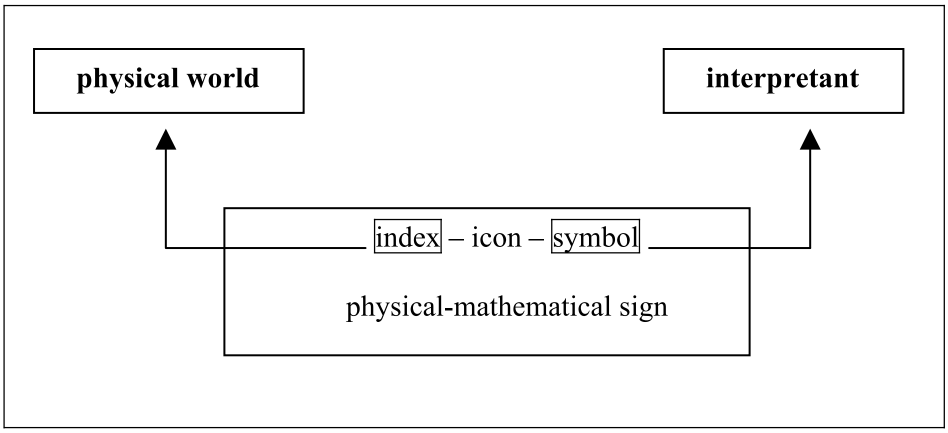


Figure 1. The semiotic triad.

However, careful attention must be paid to the fact that the triad *object/sign/interpretant* must be understood as being historically contextualized. If we do not consider it in this way, we risk being misled by historical and philosophical ingenuity, particularly so far as the relation between sign and object is concerned. Nothing, neither the sign, object, nor the interpretant are a-historical and a-cultural entities.

It is worthwhile noting here, that although we do not entirely concur with Peirce's proposal, we believe that the triadic approach which has just been outlined, is extremely efficient in clarifying relevant aspects of the contemporary physics.

3.2 The physical-mathematical sign as an icon

The icon is a representation of the object and therefore it is something similar to the object, where the similarity is intended to be between the relations of the representations and those of the elements with which the object is made. This similarity must be considered as a conjecture, as it follows from the fallibility of its constructor: man.

Apart from being a representation, an important characteristic of the icon is its independence. In fact, once constructed for a given purpose, it may be manipulated as an entity in itself, or, in other words, as having an interpretant without any connection to any object.

The most classical way of working within a physical-mathematical sign, thought of as an icon, is that concerning its logical organization. This may occur at different levels of abstraction. For example, we can work like A. H. Lorentz who, in his *Theory of Electrons*, reformulates the classical electrodynamics as a theory of principles (Lorentz, 1915), to use an Einsteinian terminology (cf. Einstein, 1934). An analogous logical reconstruction is proposed by H. Hertz in his classical *Die Prinzipien der Mechanik* (Hertz, 1894).

One immediately perceives that the logical reorganization of a physical-mathematical icon in terms of a theory of principles is nothing other than the most intuitive physical phase of the axiomatisation. The more axiomatisation is abstract, the more the physical significance of what is axiomatized recedes. From this point of view, the axiomatisations of quantum mechanics proposed by P.A.M. Dirac (Dirac, 1958), J. von Neumann (von Neumann, 1932) and by G.W. Mackey (Mackey, 1963), are totally different: the former is the least abstract one, the latter is the most abstract one.

But even attempts for the unification of more icons in one, which comprises all of them, or reducing more icons to a fundamental one, are efforts which have to do precisely with the iconic aspect of the physical-mathematical sign.

3.3 The physical-mathematical sign as an index

We have said that the physical-mathematical sign as an icon is a conjectural representation that yearns, in one way of another, to be similar to the physical world that it represents, but to have autonomy once it is constructed. The physical-mathematical sign as an index is also involved with the physical world, but this relation is much more narrow.

An index is something that indicates. It thus follows that the physical-mathematical sign as an index has an intentional value, or in other words, it indicates the physical world that, as an icon, it represents. Actually, the

relation between sign and object is twofold. On the one hand, it moves from the sign to the object (this is the gnoseologic aspect) and on the other hand, from the object it moves to the sign and then, through this, to the interpretant (this is the methodological aspect of the empirical control). This twofold relation is feasible only because the physical-mathematical sign, as an index, indicates something beyond itself. In fact, to indicate comprises two equally essential and important moments:

1) the assertion of the existence of something beyond the physical-mathematical sign;
2) the possibility of checking whether the indication is indeed correct, that is, whether what the physical-mathematical sign indicates, is truly correct.

For example, quantum field theory as an icon indicates certain elementary particles, and as an index it tells us that, apart from the formalization which represents them, they should indeed exist. It thus follows that to indicate means, on the one hand, the affirmation of the presumed existence of certain particles, and on the other hand, the possibility of checking whether this assumption is indeed founded.

It is worthwhile noting here, that there are mathematical elements within a physical theory which have no referent (for instance, Dirac's δ), or which have a dubious referent (for instance, the Higgs boson). There are also mathematical elements whose referent is closely linked to the interpretant as a whole (for instance, the wave function of quantum mechanics).

Another aspect that should be kept in mind regarding the physical-mathematical sign as an index, is that it also has a very relevant relation with the interpreter. In fact, at the moment in which the interpreter constructs the physical-mathematical sign, correspondence laws are contemporaneously posed such that they provide the sign with a given physical significance.

An example of this type of correspondence laws may be found in any textbook concerning quantum mechanics, where you find the three usual statements that give a physical meaning to the entities of an Hilbert space.[43]

Therefore, as of the beginning, the physical-mathematical sign indicates something and it is precisely this intentional aspect that allows it, as an icon, to represent (albeit conjecturally) the physical world. In addition, it is in this indicative aspect of the sign that lies the reason for its potentiality to

[43] We are speaking of the following three statements: 1) Every state of a physical system s, at the time of t, is described by a ket $|\psi>$ of the Hilbert space H. 2) Every measurable physical quantity A of a quantum system is described by an operator A defined in H and this operator is an observable. 3) The results of the measurements of a physical quantity A are given by the eigenvalues of the corresponding observable A.

discover new physical entities. And this discover may occur according to two modalities:

1) *By attributing physical significance to mathematical entities in order to "make ends meet"*. Very often when working with the physical-mathematical sign as an icon, new mathematical entities must be inserted This is either in order to eliminate a theoretical contradiction, or to unite two theoretical possibilities, or to ensure that the theory is able to coopt an otherwise negative experimental result. But the mathematical sign is also an index and thus, when inserting these new entities, there is the possibility these may indicate something other than themselves. In other words, they may disclose something in the physical world that could not have been denoted beforehand. For example, in 1892, A.H. Lorentz introduced a new system for the transformation of coordinates in order to resolve the aporia by which classical mechanics was invariant for Galilean transformations, whilst it was not for classical electromagnetism. At the beginning, the new transformations were considered as mathematical tools only, but in 1895, Lorentz assigned a physical significance to them. In our language, he first worked inside the icon in order to solve a formal problem by proposing an *ad hoc* correction. Then he began to consider this new "piece" of the physical-mathematical sign as an index also, and thus as something which indicated a real physical entity.
2) *By attributing physical significance to mathematical deductions.* We have seen that the icon has an autonomous life and this is true particularly from the deductive point of view. Once the icon is constructed, it is possible to manipulate it by deductively extracting results which may not have been contemplated at all at the time of its construction. Knowing that the sign is not only an icon but also an index, what is deduced also indicates something beyond itself, something that should exist in the physical world, thereby enticing one to determine whether this something really exists. This, for example, is what occurred with the deductive consequences obtained from general relativity after it was proposed.

3.4 The physical-mathematical sign as a symbol

To consider the physical-mathematical sign as a symbol also means thinking about how and why it is constructed. There certainly is no constraint in constructing the physical-mathematical sign, but this freedom is definitely not arbitrary since the creation of the sign is linked to both the historical context in which one works and to the problem one would like the

sign to solve. The historical constraint is so obvious that it appears banal: it is impossible to use mathematics that does not pertain to the time in which the mathematical sign is constructed. Newton would not have been able to write his *Principia Mathematica Philosophiae Naturalis* with the variational method because it did not yet exist; Maxwell did not write classical electrodynamics with the tensorial formalism, since this still had to be.

Actually, most of the physical theories are constructed by what may be called *prefabricated mathematics*, that is, a mathematics already existing "in the market". Metaphorically, one may think of a physicist like a person who goes to the market of mathematics to take what he needs to construct his theory. Einstein took Ricci and Levi-Civita's tensorial calculation for general relativity; for his work on quantum mechanics, von Neumann took Hilbert space; Weyl and Wigner took Lie groups for their works on physical symmetries. All these physicists used a prefabricated mathematics in the sense that it was constructed before of the physics in which it was then utilized[44].

However, a physicist does not necessarily have to avail of prefabricated mathematics, since he himself may create the mathematics he requires. This is Dirac's case when he introduced the pseudo-function δ to solve the problem of the continuous spectrum in quantum mechanics, or Feynman's case with his diagrams that not only visualized, but also formalized the interactions in quantum electrodynamics, as Dyson showed afterwards.

Thus on the one hand, there is the use of prefabricated mathematics, and on the other, the *ad hoc* creation of mathematical tools for the topic being dealt with, which may not necessarily be logically well-done.

[44] There actually are cases in which mathematics was available as of centuries without anyone realizing its existence. This is the emblematic case of the theory of the conic sections proposed by Apollonius of Perga in the III century B.C and used by Kepler in the XVII century A.D. Precisely the fact that even hundreds of years after it was created, a certain type of mathematics is used to create the physical-mathematical sign, brings one to the conclusion that it is impossible to know beforehand whether the given mathematics will be useful for theoretical physics. An evident example of this impossibility of knowing in advance what will be and whether there will be use for a certain type of mathematics is recalled by Dyson (Dyson, 1964). He relates the dispute between O. Veblen and J. Jeans in 1910 concerning the amendments to the mathematics syllabus at Princeton University. So, regarding which mathematics should no longer be taught, Jeans proposed eliminating group theories "as a subject which will never be used in physics"! The fact that it is impossible to know beforehand which mathematics will be used within a physical theory has two consequences. The first one is that it is impossible to distinguish between pure mathematics and applied mathematics. In fact it is obvious, as Browder maintains, there is no certainty that mathematics which has no application today, may not be applied tomorrow (Browder, 1976). The second consequence is that the more mathematics a physicist knows, the more possibilities he has to construct a good physics.

The problem now arises as to which mathematics should be utilized to construct the physical-mathematical sign. Actually, we do not have only one problem, but two:

1) the problem regarding the fact that using a particular mathematics may, at times, entail dealing with the philosophical interpretation with which that mathematics was linked up until that moment;
2) the problems related to the facts that the same physical situation may entail several different physical-mathematical signs, and that the same physical-mathematical sign may be constructed by starting from more than one mathematics.

Let us begin with the first problem. To utilize one type of mathematics rather than another may sometimes mean having to deal with a particular interpretation of the physical world. An example of this is given by Schrödinger's utilization of differential equations in his formalization of quantum mechanics. This method also involved an attempt to insert quantum mechanics within a predefined philosophical conception. This was founded on the fact that from Cauchy onwards, the formalized aspect of determinism was based on Cauchy's theorem of the existence and uniqueness of solutions of the differential equations. Totally different from a conceptual point of view, is Heisenberg's utilization of the matrix theory that was meant also as a renounce of Schrödinger's classical ideas.

So far as the second point is concerned, i.e., that we may have more than one physical-mathematical sign for the same physical situation and more than one mathematics for the same sign, one faces an epistemological problem which has its contemporary formulation in the second half of the XIX century.[45] In fact, it was Hertz in his 'Introduction' to *Die Prinzipien der Mechanik* who first in an explicit way stated that the same physical situation might be formalized by many theories. In other words, *there is an underdetermination of the physical-mathematical sign by data*. That is, the empirical world cannot settle the question of which is the best physical-mathematical that can represent it.

However, there is not only the underdetermination of the physical-mathematical sign by data, but also the *underdetermination of mathematics by physical-mathematical signs,* since, as has already been said, there is more than one mathematics which may be utilized to construct it. For example, and apart from the problems connected with the philosophical interpretation associated with the mathematics utilized, we may adopt the theory of differential equations or the matrix theory to construct quantum mechanics.

[45] Actually its original formulation has to be dated back to the Hellenistic astronomy.

3.5 The complementarity of the three aspects

Speaking of the physical-mathematical sign as an index, an icon and as a symbol, we have shown how each aspect is linked to the other two. Icon, index and symbol are not three totally independent aspects of the sign, but they are three moments which must be considered contemporaneously in order to fully comprehend it. To neglect one, or to favour another means obscuring the integral significance of the sign.

If we considered the physical-mathematical sign as an icon only, it would mean overlooking the fact that it also indicates something beyond itself, thus yielding to instrumentalism, or to empty formalism. In this case, the physical-mathematical sign might become only a formal abstract game that would be increasingly farther away from the physical referent. But in this manner, *one would risk debasing physics, which is instead a form of knowledge concerning the empirical world.* On the other hand, this is the risk one faces if one wishes to overcome theoretical difficulties such as, for example, those met in the quantisation of the gravitational field. Obviously, these difficulties impel the pursuit of abstract mathematics, but when the indical aspect is taken into account, they may be a source for important discoveries which were unthought of before, as we will see when we speak of *Dirac's methodological revolution*. But, we would like to emphasize it, *such an indical aspect has to be taken into extremely great consideration*.

However, if we considered the physical-mathematical sign as an index only and neglected it is also a conjectural icon, i.e., a tentative representation, and as a symbol, that is, a man's product which is historically contextualized, we would succumb to naive realism. The sign, with its correlated interpretant, would be no longer seen as a formalized attempt to capture the physical world, but as its precise mirror image. Everything would be considered only as an index. *In this manner, we would forget the historicity of the physical knowledge and its hypothetical status that is connected with the fallibility of the physicist.*

Lastly, if we considered the physical-mathematical sign as a symbol only, we might succumb to far-fetched conventionalism, or to more radical constructivism. In the former case, the interpretant would be considered as the result of an agreement; a negotiation between scientists in which the real world has no active role. In the latter case, the interpretant would be everything; it would be what creates the real world, which cognitively would exist only because the interpretant, formalized by the sign, exists. From this point of view, quarks would exist only because the standard model exists. All this would be the creation of the physicist who, were he God, at the time in which he constructs the mathematical sign and the correlated interpretant, creates the external world.

4. A FALSE PROBLEM: "WHY IS MATHEMATICS EFFECTIVE IN PHYSICAL SCIENCES?"

Let us now proceed to the main question of the first part of this essay: "Why is mathematics effective in physical sciences?". Let us ask ourselves: are we truly facing a real problem?

If we were to take mathematics in itself, that is, the formal part of a physical theory, it would indeed appear there was something strange and enigmatic in its effectiveness. But, as we have tried to illustrate, *the physical theory is not something to which mathematics might be added externally, thereby asking ourselves the reason for this effectiveness. The modern and the contemporary physical theories are physical-mathematical signs. They are something that cannot be divided into a mathematical part and non-mathematical part*.

The real problem does not lie in asking ourselves the motive according to which mathematics is effective, but, ultimately, in questioning the reason why physics in its entire iconic, indical and symbolic aspects is effective. Therefore, *to pose the problem of the effectiveness of mathematics in physical sciences is to pose a false-problem, that is, a problem that does not exist since mathematics is an indivisible part of the modern and contemporary physical theories.*

A mathematics that is effective in capturing the physical world does not exist other than as a mathematics with which the physical theory is constructed. Hence, it is the physical-mathematical sign as a whole that has to be considered as effective. However, in this manner, the problem shifts and becomes that of the effectiveness of the physical-mathematical sign. In other words, it becomes the problem of the effectiveness of human knowledge, since a physical-mathematical sign is one of the ways through which human knowledge acts. But this is a completely different problem and we do not deal with it.[46]

5. DIRAC'S METHODOLOGICAL REVOLUTION

In 1931 Dirac wrote (Dirac, 1931):

[46] Note that we are claiming that the problem of the effectiveness of mathematics in physical sciences is a false-problem since we cannot have contemporary physical sciences without mathematics. We are not claiming that metaphysical problems connected, for example, with a possible Platonic interpretation of the rooting of mathematics in the world is a false-problem. We do not tackle this question, as said in the introduction.

> There are at present fundamental problems in theoretical physics awaiting solution, *e.g.*, the relativistic formulation of quantum mechanics and the nature of atomic nuclei (to be followed by more difficult ones such as the problem of life), the solution of which problems will presumably require a more drastic revision of our fundamental concepts that any that have gone before. Quite likely these changes will be beyond the power of human intelligence to get the necessary new ideas by direct attempts to formulate the experimental data in mathematical terms. The theoretical worker in the future will therefore have to proceed in a more indirect way. *The most powerful method of advance that can be suggested at present is to employ all resources of pure mathematics in attempts to perfect and generalise the mathematical formalism that forms the existing basis of theoretical physics, and after each success in this direction, to try to interpret the new mathematical features in terms of physical entities.* (p. 60, our italic).

This is a real genuine revolutionary change in methodology; and only a few, even amongst philosophers of science and historians of science, are apparently aware of it. With those words, Dirac emphasized that the relation between mathematics, the physical-mathematical sign and the physical world has to be considered under a different light. Let us see how.

We began our essay by recalling how mathematics became involved with physics when the numbers, obtained from the measuring instruments, started to be considered by the philosophers of nature. In this way, it began the path which brought to contemporary physics; in other words, to a discipline in which distinguishing mathematical and physical components is senseless, but where we have a physical-mathematical sign constructed by using mathematics.

From the history of physics, we know that this long course has a topic moment when the laws describing terrestrial phenomena began to be written in a mathematical language rather than in a natural one.

Galilei proffers a paradigmatic example when he asked himself which was the law that gave significance to the phenomena of the falling bodies. He answered in mathematical terms, with a phenomenological law that allowed him not only to precisely describe but also to predict the evolution of those phenomena.

From then on, and for a prolonged period of time, this was the procedure: by reflecting on the observed and measured physical phenomena, one looks for a phenomenological law written in mathematical form that, conjecturally, represents them. In other words, starting by reflecting on what happens in the empirical world, one tries to conjecturally find the correct physical-mathematical sign.

If Galilei proceeded from the reflection on the observed evolving phenomena to the tentative phenomenological laws, Newton did something more abstract: from the reflection both on phenomena and on the phenomenological laws he knew, he passed, always conjecturally, to the laws from which the previous ones could be deduced. This is precisely the Galilean-Newtonian method. It consists into two steps: a) the Galilean step, dedicated to the formulation of the phenomenological laws describing the evolution of the phenomena observed in the empirical world; b) the more abstract Newtonian step, aimed at the formulation of general laws from which what obtained in the first step could be derived.

At this point, we think that it is useful to introduce a tripartition among physical-mathematical signs, even if here we do not discuss it profoundly. In particular, they may be classified as follows:

1) the *evolutive laws*, (they include also the phenomenological laws of the Galilean phase) which are the laws closer to the empirical world since they describe the temporal evolution of the phenomena and therefore they permit us to represent what happens in a given place and in a given time (for example, the law of motion, the wave function solution of Schrödinger's equation, etc.). Of course, these first level laws may be checked directly by experience[7].
2) the *frame laws* (generally discovered in the Newtonian phase), which enable the deduction of the *evolutive laws* as their solutions (for example, the equations of classical dynamics which enable the inference of the law of the falling bodies and, in general, the classical laws of motion; Maxwell's equations; the Schrödinger's equation; etc.). In this case, their empirical control is realized *via* theoretical *modus tollens* because of the results obtained at the level of the *evolutive laws*;
3) the *principles* which, in a certain sense, mark the boundaries of the working domain of the frame laws, and at times, also promote their construction. Here, we are thinking about principles such as that of causality which often have a *metaphysical* counterpart, but we are also thinking about *formal* principles such as those of symmetry from which, for example, owing to the powerful Noether's theorem, we can derive the frame laws of conservation.

We think that the role of time is extremely relevant in characterizing the three classes of laws. While in the first category time plays an explicit and

[7] It could be objected that there are also the so-called co-existence laws (for example, the Ohm law), that is, the laws which seem to be time-independent. Actually, the time-dependence can always be found through a deeper analysis.

important role, this is not the case for the other two categories: the frame laws and the principles refer to something invariant in time. In a certain sense, they refer to something that does not occur, but it is. Thus the knowledge offered by the frame laws and principles refers to something which does not depend on time; something atemporal. However from the atemporal representations given by such frame laws and principles, we can obtain, *via* deduction, the temporal representation, offered by the evolutive laws, of the phenomenic world.

It is interesting to note that the final outcome of the Galileian-Newtonian method concisely illustrated above was highlighted also by A. Einstein in a renowned letter dated 7 May 1952 to his friend Solovine (Fig. 2).

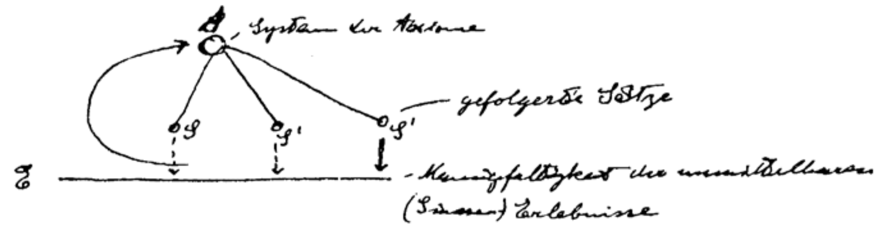
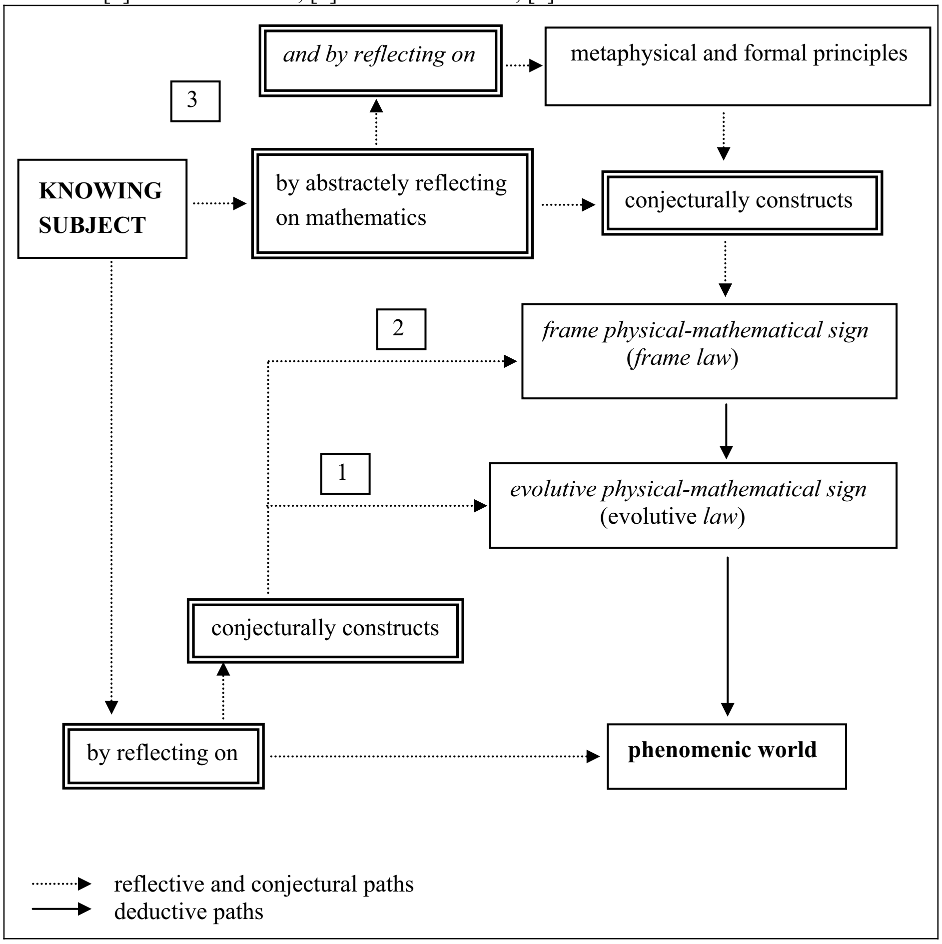


Figure 2. Einstein speaks about *System der Axiome* (our frame law, or frame physical-mathematics sign), about *gefolgerte Sätze* (our evolutive law, or evolutivephysical-mathematical sign), and about *Manningfaltigkeit der unmittelbaren (Sinnes) Eerlebnisse* (our phenomenic world). For a discussion of his letter, cf. (Miller, 1986).

But at a certain point, something changes, as Dirac wrote in the quoted passage: "it will be beyond the power of human intelligence to get the necessary new ideas by direct attempts to formulate the experiment data in mathematical terms". Therefore, whilst working within certain mathematics, at a certain point one may realize that it may be a physical-mathematical sign. In other words, one may realize that the mathematics with which is working also has an indical aspect that indicates something which exists in the world. This is what Dirac meant. In this case, it is by reflecting on the pure mathematics that we have a direct access to the level of the frame laws. In this manner, the mathematical thought assumes the role of a powerful creator of conceivably new physical-mathematical signs, and thus, as a promoter of the discovery of new phenomenic aspects of the world, derivable from those framing laws (Table 1.)

Table 1. [1] Galilei's method; [2] Newton's method; [3] Dirac's method



Dirac himself used this method when, in 1928, he adopted the spinor geometry in order to formulate his relativistic equation for the motion of the electrons.[8] This equation provided the prevision of the existence of the

[8] Even if Einstein, due to the letter to his friend Solovine, might be indicated as an emblematic example of those who followed Galilei-Newton's method, some historians of physics identify him as the one of the forerunners of what we call Dirac's method. In fact, it is sufficient to recall that it was his reflection of the principles of symmetry (covariance of the equations of motion with respect to Lorentz and general transformations) which enabled him to formulate both the frame laws of special relativity and those of general relativity, which amongst other things, allowed the prediction of nuclear energy, gravitational lenses, and black holes. In short, we are not suggesting that Dirac was the first who proposed the new method, but only that he was the first who put it in a clear and explicit manner.

positron, or in other terms, of the antimatter, which had neither been seen nor conceived of before and was discovered only two years afterwards by Anderson.

This role of the mathematical thought as a machinery to construct physical-mathematical signs and, therefore, as an inductor of discoveries of new phenomenic aspects that enhance and develop our knowledge of the world, is certainly the most relevant point of *Dirac's methodological revolution*. In this case, the physical-mathematical signs are something mentally and hypothetically constructed at the level of the *frame laws* and *principles*, simply following criteria both of internal coherence (among the propositions from which they are made up) and external coherence (with the other physical-mathematical signs).

It appears evident that the frame laws are much richer of cognitive content than the evolutive laws. In fact each one of the former implies the potential knowledge of innumerable evolutive laws of possible phenomena. It follows that discovering frame laws means enlarging our capability of knowing the world in which we live. It so happened for example with Dirac's equation that permitted us to foresee antimatter, which is now conceived as one of the pieces of the furniture of our world. And the same may be affirmed for the discovery of the curvature of space and time, now clearly visible, as gravitational lenses, in the distant clusters of galaxies.

This way of arriving to new knowledge through purely abstract thought reminds that of ancient philosophy which brought to the formulation of hypothetical metaphysical propositions, with the difference that: while the latter had to receive the general consensus through an act of faith, the hypothetical frame laws, formulated through mathematical thinking, can be empirically checked in the world.

The methodological revolution perceived and clearly illustrated by Dirac is the revolution that brings us to a substantial portion of contemporary theoretical physics[9], which favors abstract mathematical thought as a way for the construction of physical-mathematical signs that will then possibly bring us to the discovery of new entities and phenomena. The potentiality of this revolution may be well illustrated by spinor geometry, as we are going to concisely show.

[9] After the XVII Century the Galilean-Newtonian method has allowed to formulate the frame laws in a great portion of physical science. It is still adopted today as research method in several sectors of condensed matter physics, elementary particle physics and astrophysics some of which are still in the first phase: that of the phenomenological laws. It is interesting to observe that if Einstein would not have discovered general relativity, today it could be possible to discover it as a frame law, with the Galilean-Newtonian method, after the discovery of gravitational lenses.

6. THE EXAMPLE OF SPINOR GEOMETRY

Spinor geometry was discovered in 1913 by the outstanding mathematician É. Cartan who emphasized its remarkable elegance. It is based on what he called “simple spinors”.[10] Cartan also suggested that the fundamental geometry of nature may be a spinor geometry rather than an Euclidean geometry (Cartan, 1937). In fact, Cartan showed how Euclidean vectors may be bilinearly constructed from spinors, of which they are, so to speak, the square roots.[11] More precisely he conjectured that simple or pure spinors may be conceived as the elementary constituents of ordinary Euclidean vectors in so far with those spinors one can construct bilinearly Euclidean null vectors and sums (or integrals) of null vectors generally give ordinary Euclidean vectors (or minimal surfaces and strings). It was a crucial conjecture which is presently being studied and expanded further to show that it may explain several of the still obscure phenomena which were recently discovered in elementary particle physics. It may then become an efficient mathematical instrument for the discovery of new frame laws along the lines indicated by Dirac’s methodological revolution. We will try illustrating this briefly, beginning with Dirac himself, who was certainly the first person to adopt spinors in physics. This was in 1928 when he proposed the equation that represented the relativistic generalization of Schrödinger’s wave equation for the electrons (which are fermions: spin 1/2-particles) in the form

$$D(x)\,\psi(x) = 0 \qquad (1)$$

where $D(x)$ is the Dirac operator and $\psi(x)$, which represents the electron wave function, is a 4-component spinor, and x indicated a 4-dimensional space-time point (that is, the coordinates x_1, x_2, x_3, x_0=ct, where c is the velocity of light and t indicated the time).

Now suppose we adopt the Dirac’s Eq. (1) to represent a particular electron phenomenon and find a certain solution $\psi(x)$. If we reinsert it in Eq. (1), we find an identity for all values of x_1, x_2, x_3, x_0=ct. This simply means that the description of this phenomenon is valid for the whole space-time, as is Dirac’s equation, from which this solution derives,

[10] They were renamed “pure spinors” by Chevalley in (Chevalley, 1954).

[11] Note that the spinor geometry is not so intuitive as Euclidean geometry. In fact the former deals with null Euclidean vectors and totally null Euclidean planes whose vectors are all null and orthogonal, which may not be visualized by our common intuition.

In order to illustrate the role of spinor geometry it is necessary to represent Dirac's equation in 4-dimensional momentum space (or space of velocities). This entails taking the Fourier transformation of Eq. (1), which becomes

$$D(p)\psi(p)=0 \tag{2}$$

where p indicates a 4-dimensional point in the momentum space (with coordinates $p_1, p_2, p_3, p_0 = E/c$, where E indicates the energy), and $D(p)=p^j\gamma_j$; where γ_j are Dirac's 4×4 matrices and repeated indices are summed. By repeating the previous procedure, we obtain an identity in the entire momentum space.

Equation (2) may be derived from Cartan's equations that geometrically define spinors. In particular, Cartan's conjecture of the elementary nature of spinor geometry may be explicitly formulated by expressing the Euclidean 4-vector p in terms of spinors as follows:

$$p_j=\varphi^\dagger\gamma_0\gamma_j\psi \qquad (j=1,2,3,0) \tag{3}$$

where φ is an arbitrary spinor and $\varphi^\dagger$ its Hermitian conjugate.

If in Eq. (2), we again substitute a particular solution $\psi(p)$, taking into account Cartan's conjecture in the form (3), we once more obtain an identity. However, this time, it is for all values of φ, or for the entire spinor space. This is a geometrical space that is different from space-time. It may be interpreted as the purely geometrical (spinorial) origin of Dirac's equation as an atemporal frame law.

In this manner, Dirac's equation represents a sound example of a physical-mathematical sign, in particular a frame law, obtained working with pure mathematics. It predicted a new evolutive law; that of antimatter, which was discovered afterwards.

It may be shown that what we briefly exposed here for Dirac's equation may be extended to the equations of motion for higher component spinors representing fermion multiplets. In the general case, one adopts again Eq. (3), where $j=1,2,...,2n$; that is, for a $2n$ dimensional momentum space, and where, if and only if, ψ is a simple or pure spinor, the vector with components p_j is null, thus well representing Cartan's conjecture. This means that the momentum spaces where to deal with the physics of the fermions will be compact: equivalent to spheres embedded in each other. Furthermore, it can be found that most of the elementary particle properties may be derived from the four division algebras existing in mathematics. These are: real and complex numbers, quaternions and octonions.

In particular complex numbers appear to be at the origin of charges: both the electric, the weak and the strong ones. Quaternions appear to be at the origin of the so-called isotopic spin symmetry SU(2) (discovered by Heisenberg), which in turn is the origin of the proton-neutron symmetry of nuclear forces and also at the origin of the electroweak model (co-discovered by Salam). Quaternions explain also the origin of other “mysterious” properties of physics, for example, the signature of space-time.[12] From this result we may conclude that, according to Cartan’s simple spinor geometry, the Minkowsky space-time of special relativity appears to be simply the image of quaternions in nature. Octonions, finally, have seven imaginary units and they may be at the origin of the so-called SU(3) internal symmetry of flavours and colours, recently discovered in elementary particles (Budinich, 2002).

Therefore, in keeping with Dirac’s indications, spinor geometry appears to be one of the few feasible mathematical instruments needed to overcome the severe difficulties that have hindered the progress of theoretical physics for several decades. The most promising idea derives from Cartan’s conjecture on the non-elementary nature of Euclidean geometry. Already now this provides a picture that displays a marked parallelism between geometry and physics.

In fact, we know that classical mechanics of macroscopic bodies is well described with the Euclidean geometry and in space-time. However, neither macroscopic matter nor, according to Cartan, Euclidean geometry, are elementary. The “elementary constituents” of the macroscopic matter are fermions and those of Euclidean geometry are, in Cartan’s view, simple spinors. Therefore it follows that the spinors are the appropriate mathematical entities for the representation of the mechanics of fermions. The resulting mechanics is wave mechanics which then should be the “elementary constituent” of classical mechanics, as in fact it is, and in which the Euclidean concept of point-event has to be necessarily substituted by that of an integral of null vectors, which happens to be the so called string and which is non local. It could appear then that, coherently with Cartan’s conjecture, it is in this framework of spinor geometry (in the space of velocities) that wave mechanic should be properly discussed and understood (rather than in Euclidean space-time geometry, which instead is well appropriate only for the description of classical mechanics of macroscopic bodies).

[12] Note that a quaternion is an hypercomplex number characterized by 1 real and 3 imaginary numbers, or vice versa.

Thus we have a new physical-mathematical sign, and we have it simply by reflecting on pure mathematics. A sign which once more, after general relativity, shows a further geometrization of physics which, in this case, is identified with an abstract and elegant form of geometry, that of Cartan's simple spinors.

7. CONCLUSION

In this essay we have tried, through semiotic approach, to illustrate the relation between physics and mathematics, showing how:

1) one must speak about the physical-mathematical sign as an entirety;
2) the problem of the effectiveness of mathematics within physics may be considered as a false-problem due to the misunderstanding of what is contemporary theoretical physics;
3) the truly relevant fact is the methodological revolution that Dirac singled out and according to which the physical-mathematical sign is no longer constructed by conjecturally reflecting on the physical phenomena, but by working with pure mathematics and endeavoring to understand whether it also has an indical aspect and therefore if it is also a physical-mathematical sign**,** as well exemplified by Cartan's spinor geometry.

Keeping the last point in mind, we hope, particularly because it is rather undervalued, that the transition between Galilei-Newton's method to Dirac's method may become a field of research for both historians of contemporary physics as well as for philosophers of science.

REFERENCES

Browder, F. E., 1976, Does pure mathematics have a relation to the sciences?, *American Scientist* **64**: 542-549.

Budinich, P., 2002, From the geometry of pure spinors with their division Algebrasto Fermion physics, *Foundations of Physics* **32**: 1347-1398.

Cartan, É., 1937, *Leçons sur la Théorie de Spineur*, Hermann, Paris.

Chevalley, C., 1954, *The Algebraic Theory of Spinors*, Columbia University Press, New York.

Dirac, P. M. A., 1931, Quantised singularities in the electromagnetic field, *Proc. Roy. Soc*. **133**.

Dirac, P. A. M., 1958, *The Principle of Quantum Mechanics*, Oxford University Press, Oxford.

Dyson, F. J., 1964, Mathematics in the physical sciences, *Scientific American* **129**: 128-145.

Einstein, A., 1934, *The World as I See It*, Covici-Friede, New York.

Hertz, H. R., 1894, *Die Prinzipien der Mechanik (In neuem Zusammenhange dargestellt)*, Johann Ambrosius Barth, Leipzig.

Koyré, A. 1948, Du monde de l' "a-peu-près" a l'univers de la precision, *Critique* **28**.

Lorentz, H. A., 1915/2004, *The Theory of Electrons*, Dover, New York.

Mackey, G. W., 1963, *Mathematical Foundations of Quantum Mechanics*, Benjamin, New York.

Miller, A. I., 1981, *Albert Einstein's Special Theory of Relativity*, Addison Wesley Reading, Mass.

Peirce, C. S., 1895, That categorical and hypothetical propositions are one in essence, with some connected matter, in *Collected Papers of Charles Sanders Peirce*, ed. by C. Hartshorne, P. Weiss and A. Burks, Harvard University Press, Cambridge (Mass.), 2.332-339 and 2.278-280.

Quine, W. V. O., 1976, Whither physical objects?, *Boston Studies in the Philosophy of Science* **36**: 497-504.

von Neumann, J., 1932, *Mathematische Grundlagen der Quantenmechanik*, English transl. Princeton University Press, Princeton 1955.

Wigner, E., 1960, The unreasonable effectiveness of mathematics in the natural sciences, *Communication on Pure and Applied Mathematics* **13**: 1-14; now in E. Wigner, *Symmetries and Reflections*, Indiana University Press, Bloomington 1967, pp. 222-237.

ALGORITHMIC REPRESENTATION OF ASTROPHYSICAL STRUCTURES

LUCIO CRIVELLARI
INAF-Osservatorio Astrofisico, Trieste, Italy and Instituto de Astrofisísica de Canarias, Tenerife, Spain

Abstract: In the first part of this paper (§1) we give a brief review of astronomical systems and discuss a unified approach to the study of their structure by means of the kinetic theory. In the second part we deal with the algorithmic representation of physical structures (§2), and consider successively (§3) the stellar atmosphere problem as an ideal benchmark for the structural iterative algorithms designed in order to get rid of strongly non-linear problems. The underlying aim is to show that, when searching for the right mathematics, it is the physics of the problem that dictates the most efficient way to its solution. As an example we present an iterative structural algorithm that is the numerical simulation of the physical processes occurring in a stellar atmosphere. In such a way the numerical algorithm not only offers a fast and reliable mathematical tool, but also constitutes a faithful representation of the structure of the physical system.

Key words: globular clusters; groups and clusters of galaxies; the stellar atmosphere problem; Boltzmann's equation; numerical modelling; iterative procedures.

1. ASTROPHYSICAL STRUCTURES

1.1 Structure and evolution of physical systems

In order to describe physical systems, we are driven in a natural way to ideally separate inside them individual components that can be the object of our measurements, and hence can be identified with physical magnitudes.

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 97–113.

Observations may suggest the existence of causal nexuses among the distinct components, so that we are led to individuate a hierarchy of interactions that can be expressed by means of mathematical relations among the corresponding magnitudes. In such a way, by moving from observations to measurements, we are able to get a quantitative description of the whole process.

For sake of the following discussion, let us briefly recall some basic definitions. Firstly, we will consider a **physical system** as *any arbitrary assemblage of objects that can be identified and quantified by means of a proper set of physical variables.* The state of a system will be determined by the set of values taken by the ensemble of magnitudes that are necessary and sufficient to yield the maximum possible amount of information in order to determine the physical properties of the system at a given instant, and to predict its evolution with time.

In "The Oxford Dictionary" one of the definitions of the word **structure** is: "The mutual relation of the constituent parts or elements of a whole as determining its peculiar nature or character, make, frame." Such a definition can be sharpened to suit our present context by taking into consideration some lines from the issue *structure* in Abbagnano (1964):

> In a specific sense, a structure is not any plot of relations, but a plot characterized by a finalistic order. [...] According to the biologist A.C. Moulyn, a structure would be "the form relevant to the function", as the function is "the structure changing with time".

Then we will call structure of a physical system the organization of the parts, interacting among them, into which a system can be ideally separated.

According to the laws of thermodynamics, *closed systems*, i.e. systems adiabatically isolated from the external world, reach eventually the state of equilibrium, which implies a *homogeneous and steady structure*. On the contrary, due to the lack of adiabatic walls, *open systems* are characterized by outward fluxes of matter and energy that reflect a state of non-equilibrium, which implies that the variables corresponding to the relevant physical magnitudes take distinct values at the different points of the system. In general these values will change with time. The existence of space gradients are at the origin of *transport phenomena*, hence of *non-local effect*.

The structure of physical systems will be shaped by the mutual interactions among their components. In some cases antagonist forces may compensate each other and drive to a *steady state configuration*, stable over characteristic time scales. The evolution of a system in non-equilibrium can be sometimes considered as the progressive unfolding of quasi-equilibrium configurations.

Often we will be able to yield a quantitative description of the structure and the space-time evolution of the system under study in terms of differential equations that express the relations among the relevant physical variables and their rate of change. In general we will derive them from the corresponding conservation laws that we have assumed at the basis of our picture of the phenomenological world. Thus the constraints imposed by the conservation laws take part in the configuration of the system's structure. Moreover the initial and boundary conditions of the corresponding differential equations, determined from the observational data, will specify the individual systems.

1.2 Astronomical systems

We present in Table 1 a list of the most outstanding astronomical systems. Although the Galaxy, as well as the other galaxies, contain an important fraction of diffuse matter in form of gas, dust and perhaps diffuse dark matter, in the spirit of this review we will consider individual stars as the basic building blocks of these stellar systems.

Table 1. List of the most outstanding astronomical systems formed by either stars or galaxies

individual stars	$0.5M_{Sun} < \mathrm{M} < 100M_{Sun}$
binary systems	
open clusters	up to hundreds of stars
globular clusters	$\sim 10^5$ stars
individual galaxies	$\sim 10^{11}$ stars
groups of galaxies	3 or more galaxies
clusters of galaxies	from 50 up to 10^3 galaxies

1.3 Structures and their interpretation

The aim of this Section is to give, by means of two illustrative examples, some hints about the mathematical tools required for the study of astrophysical structures.

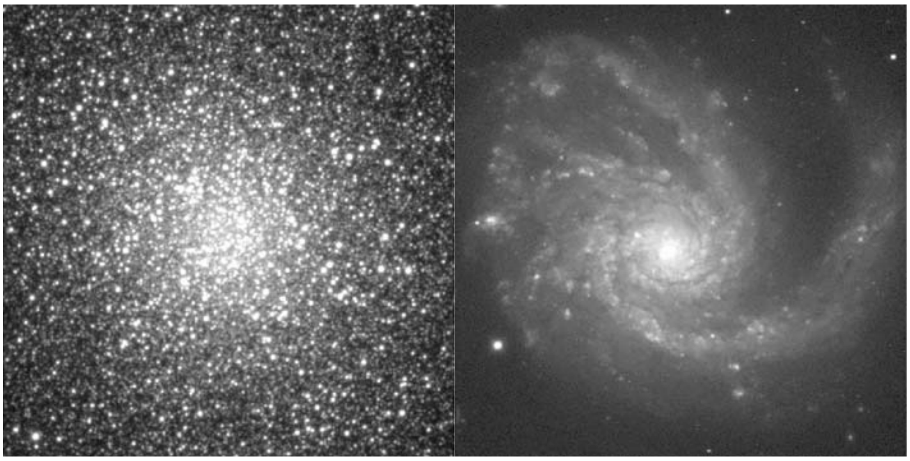


Figure 1. Left panel: The globular cluster Messier 22 (NGC 6656) in Sagittarius. (Keel, 2001) Right panel: The spiral galaxy Messier (NGC 4254) (the same source). These two images show very well the visually striking features of the selected objects, and reveal the existence of completely different structures.

Globular clusters are self-graviting collections of typically 10^5 old, low mass stars. They are characterised by a high central stellar density, and tend to have a spherical shape. By means of quantitative studies it is possible to describe the structure of a globular cluster in terms of the special distribution and the random velocities of its stars. The geometrical shape and its evolution with time are expected to be consequences of Newton's laws of motion and the gravitation as applied to an isolated spherical system composed of a very large number of point-like objects. In particular the rate of change of the stars' velocity distribution is interpreted as the result of dynamical friction, brought about by random encounters between pairs of stars.

Seminal works in this field have been carried on in the 1940's by Chandrasenkhar and Spitzer jr., whose books on the subject are still an essential reference (Chandrasenkhar, 1942; Spitzer jr., 1987). The connection of stellar dynamics with statistical mechanics is self-evident, because of the huge number of stars in the system. The specific approach to stellar dynamics in globular clusters is mainly centred on the question of the

time of relaxation of the system as the result of random stellar encounters described in terms of the classical two-body problem, and around Liouville's theorem and the solution of the equation of continuity.

Spiral galaxies are flattened systems formed by stars, gas and dust that rotate in a nearly circular fashion. The stellar component of spiral galaxies is constituted by stars with a wide range of ages that have different locations inside the overall structure. The youngest stars concentrate in a flattened disk, while older stars occupy a slightly thicker disk. The oldest stars tend to reside in a more three-dimensional distribution, which may include a central bulge, a system of globular clusters, and a low-luminosity halo.

The study of differentially rotating disks is a complex dynamical problem. Simplified analyses (like, e.g., the Wentzel-Brillouin-Kramers approximation) cannot give a complete picture of disk dynamics. In particular, there are no analytical methods that can determine the stability of a general galactic disk against arbitrary perturbations; hence N-body simulations become necessary (See Binney and Tremaine, 1997).

Whirling arms are the most striking features of the spiral galaxies. The nature of these structures, that can dominate the internal dynamics and evolution of the galaxy, is explained to day by the quasi-stationary spiral structure hypothesis formulated by Lin and Shu (1964): spirals form when compression waves propagate through the disk, growing in length and amplitude because of self-gravitational forces.

1.4 A unifying approach: the role of kinetic theory

Because of the enormous range of values taken by their fundamental parameters (i.e., the total mass of the system, its linear dimensions and the number of its members) the systems listed in Table 1 are extremely different. Nevertheless all of them share two basic features:

i) they are gravitationally bound assemblages of a very large number of interacting compact objects;

ii) the linear dimensions of the compact objects are completely negligible, compared with their average mutual distances.

These considerations justify the representation of these systems by means of the simplifying picture of a swarm of point-like particles. The paradigm of such idealized systems is yield by the theoretical model of a perfect gas, as considered by the kinetic theory.

Therefore we may describe the structure of our astronomical systems by means of the space and velocity distribution functions of their constituent particles. The evolution of the system will then be governed by the transport Boltzmann's equation for the corresponding distribution functions. Thus the mathematical methods of the kinetic theory may become, in a natural way, a

unifying approach to the study of the astrophysical structures under consideration.

The general form of Boltzmann's equation that masters the space-time evolution of the distribution function $f(\mathbf{r},\mathbf{v},t)$ can be written as

$$\frac{d}{dt}f(\mathbf{r},\mathbf{v},t)=\left(\frac{\delta f}{\delta t}\right)_{coll}. \tag{1}$$

The LHS of Eq. (1) is the total derivative of $f(\mathbf{r},\mathbf{v},t)$; the RHS denotes the collisional operator that describes the interactions among the constituent particles of the system. The specific physical conditions of each system will lead us to introduce different models in order to describe schematically the random encounters among the particles. We may try then to classify the systems according to the different approximations introduced for the corresponding operators.

1.4.1 Collisionless systems

The gravitational interaction is a long-range force. Consequently the net pull felt by any star in a galaxy (or by a galaxy in a cluster) will arise from the overall mass distribution of the system rather than from the presence of neighbouring objects. Hence the granularity of the self-gravitating matter can be ignored, and the gravitational potential Φ_s can be assumed to be smooth.

For such systems, in which random binary encounters can be fairly neglected, the collisional operator can be set equal to zero:

$$\frac{df}{dt}=\frac{\partial f}{\partial t}+\mathbf{v}\cdot\nabla f-\nabla\Phi_s\cdot\frac{\partial f}{\partial\mathbf{v}}=0, \tag{2}$$

Equation (2) is the collisionless Boltzmann equation (CBE; also called Vaslov equation), which represents a special case of Liouville's theorem.

In order to get rid of the difficulties intrinsic to the complete solution of the CBE, valuable insights can be obtained by taking the moments of the CBE. In such a way we can derive the Jeans equations (three partial differential equations for the spatial density of stars and the velocity moments) and, in a further step, to convert the latter into a single tensor equation (a form of the virial theorem), relating global properties of the system, such as its total kinetic energy and mean-square streaming velocity. The use of the virial theorem (either in tensor or scalar form) makes it

possible, for instance, to evaluate the virial mass and the mass-to-light ratio of spherical systems.

1.4.2 Binary collisions

The CBE is not valid, however, for arbitrary long time intervals. Individual stellar encounters gradually perturb stars away from the trajectories they would have taken if the distribution of the self-gravitating matter were smooth. From the comparison of the lifetime of the system with its relaxation time, defined as the characteristic time over which a star losses memory of its initial orbit, it is immediate to ascertain case by case whether the collisionless approximation holds valid or not.

For instance, the lifetime of globular clusters results to be about one hundred times larger than the relaxation time. Encounters must then be taken into account. A major difficulty arises from the fact that the value of the collisional term depends in general on the unknown distribution functions. However, if most of the stellar scattering is due to "weak" encounters, it results possible to derive a simplified form of the collisional operator by employing the Fokker-Planck approximation.

2. ALGORITHMIC REPRESENTATION

2.1 Modelling the physical world

We can ideally dissect a system into many simpler interacting parts, in order to describe its global behaviour in terms of the laws governing the elementary components. Such a process of dissection allows us to get eventually a *model* of the physical system. Modelling is an unparalleled tool for scientific inquiry because, for its own analytical nature, it can be easily translated into a set of equations, i.e. into a *mathematical model*.

A premise to this Section, that deals with the "gentle art of modelling", might be Fourier's claim (Fourier, 1830) that the relations among the mathematical functions of the physical variables and their derivatives do not just pertain to the abstract realm of Calculus; they do actually exist in the natural phenomena themselves. According to this view, the general scheme required in order to convert the mathematical models into numerical information by means of algorithms should also partakes of Nature. Although Fourier's words are the expression of a naive form of realism, by echoing the famous statement by Galileo that the Great Book of Nature is

written with geometrical characters, they suggest however that Nature displays herself to human mind through algorithms.

The steps required for moving from the physical word to its quantitative representation by means of numerical algorithms are pictured in Figure 2. The realization of these steps is the result of the efforts done by the mathematicians, especially in the second half of the last century, in order to find out a precise link between the mathematical models and the numerical information necessary to describe the physical world.

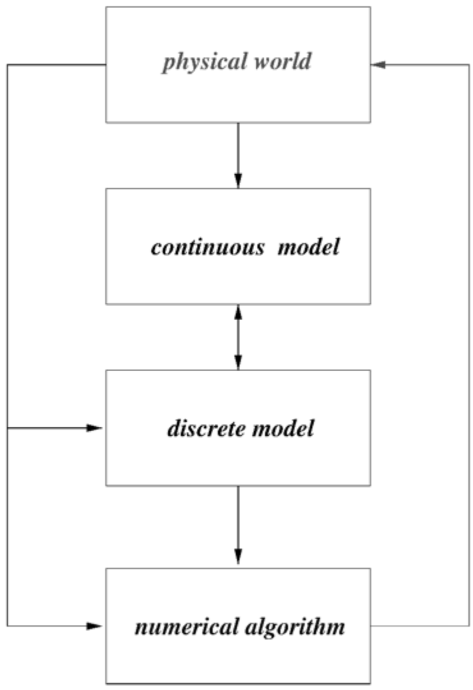


Figure 2. From the observation of the physical world to the algorithmic representation of a physical system.

2.1.1 From the physical world to the continuous model

The interactions among the different parts of the physical system and its space-time evolution are necessarily simplified and idealized in terms of a system of differential equations (either ODE or PDE), with the corresponding initial and boundary conditions.

2.1.2 From the continuous to the discrete model

In the previous step the structure of the physical system has been formalized in terms of a system of equations. But for very special cases, an exact solution will not be feasible, hence the need of a numerical solution via discretization (e.g., by means of discrete ordinate methods). In order to achieve the optimal discretization, it is not just matter of the analysis of the structure of a formula merely mathematical in character. The physics of the problem will dictate the grid of discrete ordinates, according to the scale heights of the intervening physical variables.

2.1.3 From the discrete model to the algorithm

After that the original system of equations has been transformed into the corresponding system of discretized equations, we are in a position to build eventually a suitable numerical algorithm for achieving their solution.

If each step of the above chain has been worked out properly, the structure of the physical system will have been reflected into the mathematical form of the relevant equations, and the structure of the ultimate resolving algorithm will be akin to that of the initial model. Then we can consider the numerical algorithm as an *image* of the original physical system, in other words, a *representation of its structure*.

2.2 The search for the best numerical algorithm

The original system of discretized differential or integral equation will be not, in general, linear. However, via a proper linearization technique, it may be converted into an equivalent system of linear algebraic equations, whose matrix will reproduce, for the nature and the collocation of its elements, the structure of the initial model.

What looks simple in principle results often, however, infeasible in the practice of actual computation. It is well known that the numerical inversion of large or ill-conditioned matrices is a nasty problem. In a seminal paper von Neumann and Goldstine (von Neumann and Goldstine, 1947) analysed the four main sources of error in numerical computation. In particular they discuss the errors introduced when “exact” arithmetic (i.e. transcendental operations) is replaced by “approximate” arithmetic (i.e. elementary operations that can be handled by a computer). No computing machine can perform all of its elementary operations rigorously and faultlessly because of the finite number of digits available. Even if we could master the other sources of error that arise from the translation of the physical model into a system of equations and its successive discretization, the problem of stability

brought about by the former unavoidable source of errors should always constitute a critical drawback for the numerical inversion of huge matrices, by means of either direct or iterative methods.

We are therefore compelled to seek for strategies alternative to the complete linearization technique. By keeping in mind Henry Poincare's words: "*La physique ne nous donne pas seulement l'occasion de résoudre des problèmes ... elle nous fait pressentir la solution*", we will ask to the physics of the problem under study to dictate the right approach to its solution. In the next section we are going to illustrate our point by means of the case study of the stellar atmosphere problem.

3. STRUCTURAL ITERATIVE ALGORITHMS

3.1 The stellar atmosphere problem

We call stellar atmosphere the outer layers of a star, in which the radiation flowing out of the stellar core acquires its spectral features. Spectroscopy shows that their constituting material is plasma at high temperatures, whose particles are gravitationally confined. On the other hand, the major observational evidence of an outward flux of radiation reveals the anisotropy of the radiation field, which is the signature of radiative transfer. Hence we can consider stellar atmospheres as an open boundary between the stellar interior and the interstellar medium. From the physical standpoint we have to deal with a system that consists of two main components: matter and the radiation field in which matter particles are embedded.

The steady-state conditions observed over very long time scales warrant the existence of a mechanical equilibrium among the external and internal forces acting on the particles, i.e. gravitation, gas pressure and radiation pressure. The interchange of energy between the two components of the systems and the absence (as a reasonable first order approximation) of energy sources in the atmosphere impose the additional constraint that the total sum of the internal energy of matter and the energy of the radiation field be constant. The resulting energy balance is intrinsically linked with the mechanical equilibrium.

The internal energy of matter determines its temperature, and consequently the values of the other thermodynamic variables. In turns, the thermodynamic state of matter determines the values of the radiative transfer coefficients. That is to say, the coupling between matter and radiation results strongly non-linear. Moreover, the energy balance is governed by radiative

transfer, i.e. by a transport process. Therefore the stellar atmosphere problem happens to be *non-local* and *highly non-linear*.

3.2 The equations of the problem

The problem of determining the structure of a stellar atmosphere is tantamount to describe the local concentration of matter and energy. According to the specific requirements, we shall make use of either a microscopical or a macroscopic description of the physical system.

In stellar atmospheres the temperature is sufficiently high and the density is sufficiently low for the matter particles to be localized wave packets whose extensions are small compared to the average inter-particle distance. Thus they can be idealized as point-like particles, which will have however also internal degrees of freedom as postulated by the old quantum theory. In parallel, the radiation component can be also treated according to the kinetic theory, if the specific intensity of the radiation field is described by means of the photons distribution function. Then the radiative transfer equation can be interpreted as the Boltzmann transport equation for the photon gas. It would be easy to demonstrate that the usual formulation of the RHS member of the latter in terms of the specific intensity and the transport coefficients corresponds, in the kinetic formulation, to a collisional operator that consists of a scattering term and a conventional Bahtnagar-Gross-Krook relaxation term.

All that makes it possible to employ the methods of the kinetic theory for the derivation of the fundamental equations of the stellar atmosphere problems. (This fact is a further example of the unifying role of the kinetic theory suggested above.) In the normal way, following the paradigm of thermodynamics and fluid dynamics, we will be able to derive the set of equations the express the relations among macroscopic quantities that are defined by the proper moments of the corresponding distribution functions.

From the analysis of its components and their mutual relations, we can deduce those elements that determine the structure of a stellar atmosphere. These elements, which shall therefore be taken into account by the corresponding mathematical model, are essentially:

i) the constraints imposed by the equilibrium hypotheses;

ii) the description of the physical state of matter, either through the equation of state if local thermodynamic equilibrium (LTE) can be assumed, or a kinetic treatment of the atomic level populations under non-LTE conditions;

iii) the transport equations (i.e. radiative transfer and - sometimes - convective transport);

iv) a proper set of initial and boundary conditions.

On this basis we will eventually derive the system of equations of the stellar atmosphere problem. A picture of these equations is given in Table 2, which may be viewed as a snapshot of the mathematical model of a stellar atmosphere.

Table 2. The equations of the stellar atmosphere problem. This table is taken from (Crivellari, 2002), where the stellar atmosphere problem and iterative methods for its solution are presented in full details.

constitutive equations:

$$\frac{\partial}{\partial t}\rho+\nabla\cdot(\rho\mathbf{v})=0$$

$$\frac{\partial}{\partial t}(\rho\mathbf{v})+\nabla\cdot(\rho\mathbf{vv})=\rho\,\mathbf{g}-\nabla P$$

state equation:

$$P=\frac{k}{m_H\,\mu}\rho T$$

energy equations:

$$\frac{\partial}{\partial t}\left[\frac{1}{2}\rho v^2+U(T)\right]+\nabla\cdot\left[\frac{1}{2}\rho v^2+U(T)\right]\mathbf{v}=\rho\mathbf{v}\cdot\mathbf{g}+\nabla\cdot(P\mathbf{v})$$

$$\frac{\partial}{\partial t}u+\nabla\cdot F=\int d\nu\oint d\Omega\left(\eta_\nu-\chi_\nu I_\nu\right)$$

radiative transfer equation:

$$\left(\frac{1}{c}\frac{\partial}{\partial t}+\mathbf{n}\cdot\nabla\right)I_\nu=\eta_\nu-\chi_\nu I_\nu$$

description of the microscopical state:

$$\mu,\eta_\nu,\chi_\nu$$

$$LTE\text{: } f(\rho,T); \qquad non\text{–}LTE\text{: } f\left(\rho,T;\{J_\nu\}\right)$$

3.3 Iterative solutions and the method of the iteration factors

The standard approach to the numerical solution of non-linear problems is the *complete linearization technique*, which consists of the linearization of all the equations together with their initial and boundary conditions. This method is iterative in character: the unknown variables are linearized around a current estimate of their values, that can be obtained either from an initial guess or be the result of the previous step of iterations.

At each step of iterations the resulting system of linear algebraic equations must be solved numerically. Although always possible in principle, when the dimensions of the system are huge (which is always the case in the stellar atmosphere problem) the direct solution becomes infeasible, as already mentioned in § 2.2. Therefore one must look for iterative methods of solution also for this instance.

The fact that actual non-linear problems require an iterative solution is the rationale for seeking case by case the optimum one. Again the best strategy is to look at the physics of the problem, instead of trying smart improvements of the existing mathematical methods. Moreover, very often the numerical simulation of the physical processes suits perfectly the capabilities of nowadays computers.

An implicit assumption of the complete linearization technique is that all the physical variables can be treated on equal footing. In our opinion, there are reasons of principles against such an "equalitarian" treatment. Indeed the different processes taking place in stellar atmosphere are characterized by very different height scales. Moreover, the strength of the coupling between the different phenomena may vary considerably case by case. On these grounds we propose an alternative sequential approach. According to the nature of their mutual interactions, the different physical processes are grouped into *elementary blocks*, so that each one of them contain the minimum amount of information necessary for the self-consistent statement of the corresponding physical problem, that can thus be considered as "atomic". The blocks are treated separately one by one. The current values of the external variables (i.e. those not pertaining to the block), are taken as input data that will not be altered inside the block. The equations of the atomic problem are solved to yield the current values of the internal variables, which constitute the output of the block.

The elementary blocks are organized into a sequential structure that can be straightforwardly translated into an iterative algorithm for the global solution of the problem. The most natural sequential organization of the elementary blocks for the stellar atmosphere problem would be the one we present in Fig. 3.

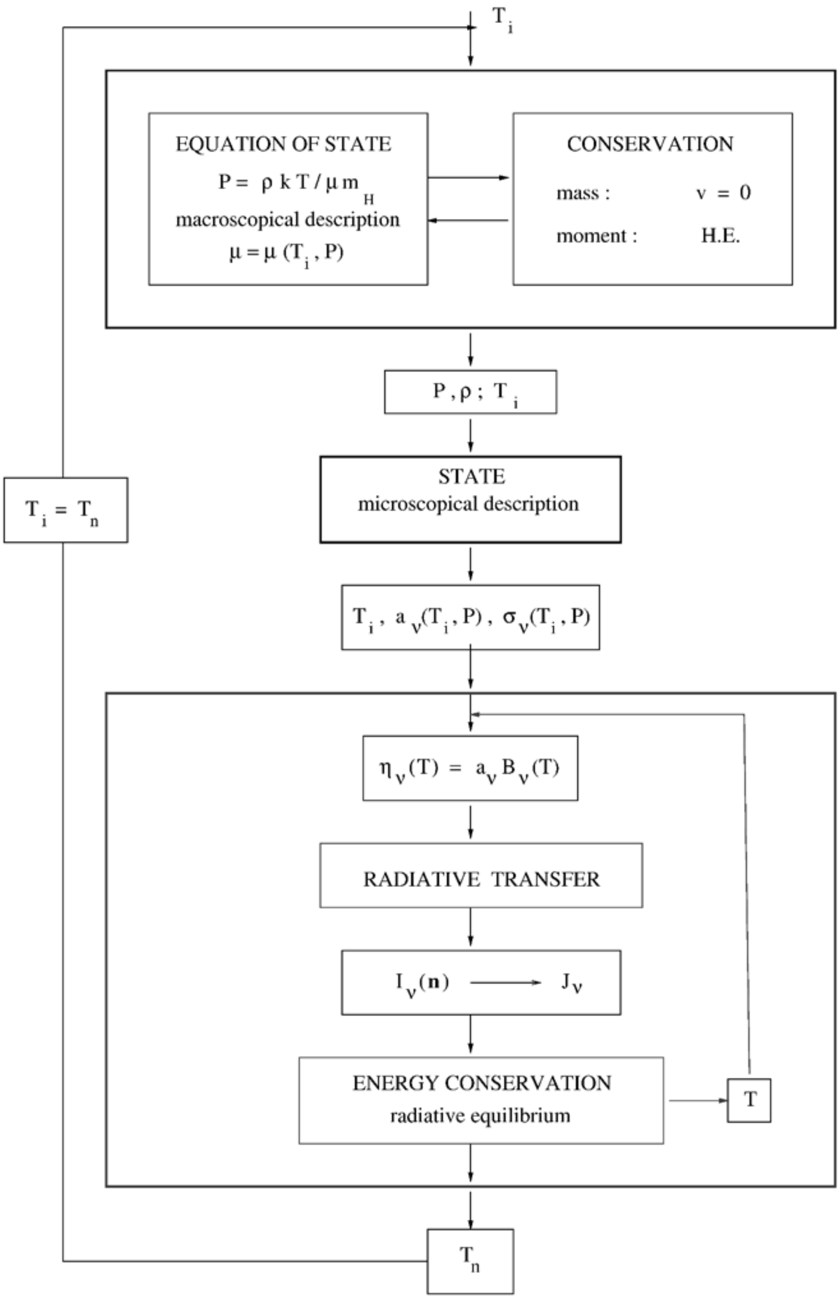


Figure 3. Flow-chart of the sequential iterative procedure.

Two *macro-blocks* can be individuated at once. The upper one, the *constitutive* macro-block, consists of the two-coupled elementary blocks that

account respectively for the mass and momentum conservation and the equation of state (here formulated in *macroscopic* terms). The lower one, the *energy* macro-block, considers sequentially the system of the radiative transfer (RT) equations and the constraint of energy conservation. In the simplified model considered here the latter condition states that the amount of radiative energy absorbed by matter must be equal to the energy gained by the radiation field through emission processes. The energy balance equation reads

$$\int a_\nu J_\nu d\nu = \int a_\nu B_\nu(T) d\nu \,,$$

where a_ν is the absorption coefficient, J_ν the mean intensity of the radiation field, and $B_\nu(T)$ Planck's function. The two macro-blocks are coupled through the equation of state, here formulated in *microscopical* terms.

The input of each step of the whole sequential procedure is the current value T_i of the temperature. The output values of the variables of the constitutive macro-block, P and ρ, are easily obtained by means of a simple iterative loop, because the physical interaction between the two elementary blocks is weak. By means of the microscopical equation of state it is possible to compute the coefficients a_ν and σ_ν of the RT equations that, together with the constitutive variables are the input data of the energy macro-block. The successive solution of the RT equations yields the current values of the specific intensity of the radiation field.

At this stage the current values of the relevant variables should fulfil the constraint of energy conservation. In general, of course, that will not be the case. Therefore we shall make use of the energy balance as an implicit transcendental equation in the unknown T in order to get a new value T_n of the temperature that satisfy the energy constraint. The whole procedure is iterated till the convergence to the correct physical solution is achieved.

The internal loop of the energy macro-block is a version of the Picard - von Neumann series. The problem is that in most of the actual astrophysical applications its rate of convergence results exceedingly slow (if the convergence to the correct solution is achieved at all) because of the very large optical depth of the medium through which photons propagate. To cut a long story short[47], the reason of the failure has to be found in the sequential treatment of the RT equations and the temperature correction. The exchange of energy between the radiation field and matter constitutes a strong physical

[47] For a detailed discussion see (Crivellari, 2002; Simonneau and Crivellari, 1999).

coupling between the two components. Consequently radiative transfer and the energy constraint cannot be treated sequentially in the energy macro-block, but the corresponding equations must be solved simultaneously.

That could be done, in principle, by including the energy balance in the source functions of the corresponding RT equations. However this straightforward strategy cannot work because the integral $\int a_\nu J_\nu d\nu$ that will appear in the source functions of the RT equations is mathematically equivalent to a scattering integral, and it is a matter of experience that diffusion problems cannot be effectively solved by means of iterative procedures for large optical depths.

The way to get out from this impasse is offered by the Method of the Iteration Factors (IFM)[48] At any step of an iterative procedure, the current values of the fundamental variables may be very far from those of the final solution. However, the ratios of certain homologous variables may be very close to their correct value, because offset errors have been mended by the division between homologous magnitudes. Such ratios (e.g., between pair of moments of a distribution function, or different kinds of average of the absorption coefficients, and so on), that result to be quasi-invariant along the run of iterations, are called *iteration factors*.

In the specific application of the IFM to the energy macro-block, inside each step of the internal loop the RT equations are firstly solved one by one in order to get the current values of the specific intensity. A set of iteration factors is successively built by taking the ratios of proper pairs of both moments of the specific intensity and averages of the absorption coefficient. By taking in the standard way the moments of the RT equation, a bolometric radiative transfer equation can be derived, whose solution will automatically fulfil the energy constraint. The coefficients of the bolometric equation will be given by the above iteration factors.

The graphical representation of the energy macro-block, given by Fig. 3, will now be altered. The single elementary block energy conservation shall be replaced by two blocks in parallel: bolometric radiative transfer and energy conservation, linked by a two-way arrow. The latter symbol denotes the channel through which information is exchanged between the two blocks by means of the iteration factors.

The above scheme quickly converges to the correct physical solution thanks to the numerical properties of the iteration factors. Ordinary sequential iterative methods carry on a burden of spurious information among elementary blocks. On the contrary, when the iteration factors are the channel through which information flows, a filtered output is transmitted as

[48] See (Simonneau and Crivellari, 1988; Crivellari and Simonneau, 1991).

input to the next blocks. That not only dramatically fastens the rate of convergence, but also warrants the stability of the iterative procedure.

ACKNOWLEDGMENT

The author heartily thanks the organizers for the invitation and the kind hospitality in Lošinj.

REFERENCES

Abbagnano, N., 1964, *Dizionario di Filosofia*, UTET, Torino.
Binney, J., and Tremaine, S., 1997, *Galactic Dynamic*, Princeton University Press, Princeton.
Chandrasenkhar, S., 1942, *Principles of Stellar Dynamics*, The University of Chicago Press, Chicago.
Crivellari, L., and Simonneau, E., 1991, The use of iteration factors for temperature correction in a stellar atmosphere when convective transport is present, *Astrophys. J.* 367: 612.
Crivellari, L., 2002, Il problema delle atmosfere stellari: un approccio strutturale; http://www.na.astro.it.
Fourier, J.B., 1830, *Analyse des Equations Déterminéés*, Paris, p. 185.
Keel, B., 2001, Keel's image collection available at: http://crux.astr.ua.edu/choosepic.html.
Lin, C.C., and Shu, F.H., 1964,
Simonneau, E. and Crivellari, L., 1988, On the application of iteration factors for temperature correction in stellar atmosphere, *Atrophys. J.* **330**: 415.
Simonneau, E. and Crivellari, L., 1999, Some iterative methods for radiative transfer problems, *Transfert de Rayonnement en Astrophysique*, Forum de GRETA, CEPADEUS Ed., Touluse, p.67.
Spitzer jr., L., 1987, *Dynamical Evolution of Globar Clusters*, Princeton University Press, Princeton.
Von Neumann, J., and Goldstine, H.H., 1947, Numerical inverting of matrices of high order, *Bulletin of the American Mathematical Society* **LIII**, p. 1022.

THE FLEXIBILITY OF MATHEMATICS

DENNIS DIEKS
Institute for the History and Foundations of Science, Utrecht University, Utrecht, The Netherlands

Abstract: Mathematics is quite unlike physics: it does not possess empirical content and lives in an independent realm of its own. It seems surprising that the partnership of these dissimilar companions, mathematics and physics, is so extremely successful. But I argue that on further reflection this success is not ‘unreasonable’: the very difference between the nature of mathematics and that of physics makes it possible for mathematics to be highly flexible and adaptable to the most diverse needs. By means of a number of examples, drawn from fundamental physics, I illustrate how mathematics, through its flexibility and versatility, achieves its great effectiveness.

Key words: effectiveness of mathematics; physics; holism; relativity; quantum mechanics.

1. INTRODUCTION

“Mathematics may be defined as the subject where we never know what we are talking about, nor whether what we are saying is true”, Bertrand Russell famously remarked (Russell, 1917). What Russell wanted to express is that mathematics is very different from empirical science: it is not about the physical world in which we live and which we can see, touch and smell. One does not have to subscribe to the details of Russell’s philosophy of mathematics to agree with this point. For example, Platonists think that mathematics *does* describe something and *can* be true (in the correspondence sense): it describes an Ideal mathematical world. Still, this mathematical heaven is completely separate from the physical world around us. Formalists, on the other hand, believe that mathematics does not describe anything at all but is a mere play with symbols, according to man-made

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 115–129.

rules. But in spite of this shift in philosophical outlook, it remains true that mathematics is without physical content.

In view of the great distance between physics and mathematics it may come as a surprise that mathematics plays such an important role in modern physics. In some instances mathematical considerations are even the dominant force in physical research. Is this proven effectiveness of mathematics in physics not hard to understand, very 'unreasonable', as Wigner put it in a famous essay (Wigner, 1960)?

Now, it seems plausible that it is not *a priori* necessary that a scientific account, in the usual sense, of the physical world is possible. The world might have been dramatically irregular, with no structural permanence at all. The concept of a 'law of nature' would not be usable in such a situation and the question of whether natural laws can be couched in mathematical language would not even arise. So, perhaps, it may be considered astounding that there is regularity in our world at all, that the concept of a law of nature actually makes sense. Perhaps one could argue that *a priori* it is more probable that there is no order than that there is—that we therefore find ourselves in an improbable situation, which justifies surprise. I am not sure about arguments of this sort—the status of the *a priori* probabilities used in them, and the justification of the values assigned to these probabilities, seem very much open to question. Moreover, Kantian or anthropic counterarguments may be defensible, about the physical conditions that have to be satisfied in order to make our own existence possible. But in this essay I will not embark on speculations about whether there might be reasons why regularities in nature exist: I am going to take the existence of such regularities as something given.

The Wignerian question then becomes: granted that there is order and structure in nature, isn't it unreasonable to expect that mathematics is highly effective? Is it not strange that mathematics not infrequently plays an inspiring role in physical research, and points the way to new results?

The answer that I want to suggest is that the very observation that mathematics has no physical content can take away most of the surprise. Indeed, exactly because mathematics is a 'freely floating construction', not tightly bound to sense experience, it is extremely flexible and versatile—and therefore useful. I will illustrate some aspects of this flexibility and versatility below, by examples from fundamental physics.

One thing I want to make clear by these examples is that the same physical situation can usually be described in a variety of mathematical ways. The mathematical toolbox is so well-stocked that researchers of different approaches and persuasions can often find a way of dealing with a subject that suits their particular tastes and enables them to pursue their own programs. Conversely, since mathematics itself is empirically empty, the

same mathematical techniques and results can often be applied to a diversity of physical situations; new insights can thus be gained at small costs by transporting old results to new contexts. The effectiveness of mathematics thus appears as a built-in feature: because of its flexible applicability anywhere where some type of order reigns, and because of its adaptability to research preferences, mathematics is likely to be effective.

2. NON-UNIQUENESS OF MATHEMATICAL MODELS

In the years during which the genesis of modern quantum theory took place, mathematical techniques from different directions were employed. Heisenberg's matrix mechanics and Schrödinger's wave mechanics, respectively, had a radically different mathematical form and fitted in with very different methodological programs. However, both formalisms were able to handle the discreteness of spectral lines, and therefore succeeded in explaining the most crucial experimental fact that classical theory could not handle. This already furnishes a first example of the flexibility of mathematics. Schrödinger, repelled by the abstract character of Heisenberg's theory, was able to find an alternative mathematical treatment that satisfied his own philosophical and aesthetic demands but made the same observable physical predictions as the abhorred rival theory. This new mathematical scheme enabled him to pursue his favourite idea, according to which quantum objects are inherently wave-like.

However, there is a limit to this kind of adaptability. One cannot impose any philosophical preference whatsoever on nature. Although mathematics is very flexible and will go a long way in meeting a researcher's wishes, it cannot guarantee that all desiderata will be implementable. Nature itself, experimental results, obviously limit the possibilities: Schrödinger was in fact unable to carry through his pet notion that particles are local spots of high density in an omnipresent continuous field. The mathematical reason for this is that the wave field is defined in configuration space rather than in ordinary three-dimensional space, which becomes important as soon as systems consisting of more than one particle are considered; an additional problem is that local regions of high field intensity will not be stable because of dispersion. These features of the wave theory prove inevitable if justice is to be done to the observed phenomena.

Nevertheless, the flexibility of the mathematical treatment permitted Schrödinger to make the most of his research program. The discrete nature of Heisenberg's calculus clearly turned out to be avoidable; a continuum

treatment could be put in its place. Mathematics afforded the maximum of flexibility compatible with empirical results.

Not long after the discussions about these issues in the nineteen twenties, von Neumann showed that both Heisenberg's and Schrödinger's theories could be seen as versions of one encompassing mathematical scheme—quantum theory as formulated in Hilbert space (von Neumann, 1932). In spite of the fact that matrix mechanics and wave mechanics are so very different—the former a calculus of discrete quantities, the latter a continuum theory—mathematics was able to provide a unifying framework. By going up one level of mathematical generality, it proved possible to transcend the seemingly unbridgeable differences and to turn the two theories into one. This exemplifies the power of mathematics in bringing out hidden similarities and common structures.

But it should be noted that this unifying power of mathematics is not directly related to effectiveness in dealing with natural phenomena. The two theories under discussion—matrix and wave mechanics—can be regarded as purely mathematical schemes. They are unified by von Neumann's Hilbert space formulation, which itself can also be seen as purely mathematical. Mathematics is able to do its unifying work here just because it is able to describe structure (in this case the—at first hidden—common structure of wave and matrix mechanics), quite independently of whether this structure represents something in physical reality. *If* some kind of structure is realized in physical reality, mathematics can be counted on to give a fitting description. This statement cannot be reversed: if mathematics defines a certain structure, we cannot count on its importance in physical theory. The unifying power of mathematics does therefore not testify to an *a priori* rapport between mathematics and physical reality.

Von Neumann's Hilbert space formulation, with its non-commuting observables, has become standard. Still, it has not remained unchallenged. The Bohm formulation of quantum mechanics does not work with Hilbert space, but with configuration space as the fundamental arena of physical processes. It operates with the classical particle concept, according to which a particle possesses a definite position and momentum at all times. By contrast, in the standard scheme physical systems cannot have both a definite value of momentum and position, because the corresponding operators do not commute.

There is no need to rehearse the mathematical details of the Bohm approach, which are well known. The point of mentioning this alternative to the standard formulation is that we have here another example of two completely different mathematical schemes that agree about the results of empirical observation. As in our previous example, it again is true that we cannot impose *everything* we might wish. In order to achieve empirical

adequacy we have to accept non-locality of interactions in the Bohm theory, for instance. But this case does provide another illustration of how mathematics allows us the maximum possible latitude in accommodating our methodological, interpretational and philosophical preferences. Mathematics thus facilitates the formulation and execution of diverse and rival research programs, and consequently enhances the chances of progress. In the case at hand, it gives us the means to investigate to what extent the classical particle concept is still viable within the quantum context. Again, granted that there is a particular structure of the world out there to be discovered (this is what we assumed to begin with), it is no miracle that the richness of mathematical tools and the corresponding variety of possible research paths help us to actually do the discovering.

3. HOLISM

The Bohm theory differs from standard quantum theory in that it operates with particles that occupy spatial points, and with many-particles systems whose state (in the sense of their location in phase space) is fixed by the states of their components—like we are used to in Newtonian physics. In this sense the Bohm theory is associated with a ‘local’ world picture. By contrast, standard quantum mechanics is ‘holistic’, because properties of composite systems can often not be reconstructed from properties of their component systems. Think, for example, of the two-electron singlet spin state, in which the total spin is definite but cannot be regarded as the sum of definite spin values of the individual particles.

The empirical results are compatible both with Bohm’s theory and with standard quantum mechanics, so they can be accommodated both within a local and a holistic treatment. More generally, discussions about ‘locality’ and ‘holism’ in physics usually cannot be decided by empirical data alone. The empirical findings have to be evaluated within a theoretical scheme—and mathematics is often able to supply adequate schemes of different kinds.

An interesting further case is furnished by electrodynamics. Classical electrodynamics has a purely local form, in the sense that the central quantities $\vec{E}$ and $\vec{B}$ are fields, defined ‘per point’. That is, an electric and a magnetic field strength are assigned to each spatial point, and these field strengths determine the force on a charged particle there. In addition, the theory works with electromagnetic potentials, ϕ and $\vec{A}$. These are also defined locally; moreover, they determine $\vec{E}$ and $\vec{B}$ via local relations. In the relativistic treatment these electromagnetic quantities are represented by the anti-symmetric electromagnetic field tensor $F_{\mu\nu}$ and the four-potential A_μ; again, both are defined locally.

The potentials are not uniquely determined by the observable phenomena: gauge transformations $A_\mu \longrightarrow A_\mu - \nabla_\mu \Lambda$, with Λ an arbitrary scalar field, change the local values of the potentials but leave the field strengths, and the forces exerted on charges, the same. In classical electrodynamics the underdetermination caused by this gauge freedom is usually considered as insignificant, because the electromagnetic potentials are regarded as purely mathematical expediencies—only $F_{\mu\nu}$ is accepted as physically real. In quantum mechanics, however, the situation becomes different: the wave function couples directly to A_μ. Even if no electromagnetic fields are present in a region (i.e., $F_{\mu\nu} = 0$), the wave function does not evolve freely if $A_\mu \neq 0$. The notorious example is the Aharonov-Bohm effect, in which an electron can move along two different paths around a solenoid. Inside the solenoid there is a magnetic field, but $\vec{E}$ and $\vec{B}$ vanish outside of it. The electron moves outside the solenoid and therefore cannot experience the fields. Still, the electron's wave function is changed because of the presence of $\vec{A}$ in the region where the electric and magnetic fields disappear.

More in detail, the wave function incurs a phase $\int_C \vec{A}.d\vec{r}$ along a path C. This phase is empirically significant: the phase difference between the two paths around the solenoid, which is given by $\oint_C \vec{A}.d\vec{r}$ with the integral taken over a closed contour surrounding the solenoid, is responsible for interference effects that can be measured.

The presence of A_μ thus has observable effects, and the potential therefore cannot simply be dismissed as physically unreal. Nevertheless, the gauge freedom $A_\mu \longrightarrow A_\mu - \nabla_\mu \Lambda$ is still there, because the integral $\oint_C \vec{A}.d\vec{r}$ is invariant under such gauge transformations. So the value of A_μ *at a point* remains unobservable; it is only the integral taken over a closed path that is measurable.

One can now choose between two positions. One is that we are dealing with a completely local theory, characterized by the real physical fields $F_{\mu\nu}$ and A_μ. It is true that the local values of A_μ cannot be observed; but according to the position under discussion this does not automatically entail that there is nothing real corresponding to A_μ. Indeed, there are many things in physics which are not directly accessible, and about which information can only be obtained in a roundabout way. It is often taken for granted nevertheless that the entities in question exist - think of atoms or elementary particles. It is natural, however, if one is convinced of their reality, to look for ways of obtaining more direct information about them. In our case, if one believes that the potentials are physically real, it is plausible to think of ways by which A_μ could be observed directly; to accommodate this theoretically, the theory should be modified. Now, suppose that such an attempt succeeds and results in a better theory, one that is able to predict more. Perhaps one is

then inclined to say: “Mathematics has miraculously led the way; even before we could measure A_μ, mathematics already indicated its existence! Mathematics is unreasonably effective.”

But we may also take the position that A_μ does *not* represent a real physical field. In that case it is plausible to look for formulations of the theory in which A_μ does not occur; one would like to be parsimonious and only represent quantities that do possess physical significance. Mathematics is an obedient servant: a formulation of electrodynamics in which the phases over closed contours (the gauge-invariant quantities that can be directly observed, as we saw above) are central can readily be found (Wu and Yang, 1975). In this formulation one starts with the ‘anholonomy’ (the mentioned phase) associated with closed curves, and there is no need to introduce local potentials. Now, suppose no evidence for the reality of potentials is ever found. One is then perhaps inclined to say: “Mathematics has miraculously led the way: even before experiments convinced us that A_μ has no physical existence, mathematics already indicated the holistic nature of electrodynamics! Mathematics is unreasonably effective.”

The moral is that mathematics is so versatile that it can fit diametrically opposed heuristics and research programs. There is no pre-established harmony between mathematics and the eventual course physics will take.

4. RELATIVITY

General relativity is sometimes adduced as an example of a situation in which a mathematical framework—differential geometry—that was developed completely independently of physical needs proved unreasonably efficient. Differential geometry was first developed as a branch of geometry by Gauss—as a metrical theory of curved two-dimensional surfaces—and then generalized to an arbitrary number of dimensions by Riemann. In the second half of the 19th century the subject underwent further evolution, through the work of mathematicians like Levi-Civita. After Einstein got acquainted with differential geometry, this branch of mathematics proved to be of decisive importance in achieving a break-through in his struggle for a relativistic theory of gravitation.

I do not think, however, that the great effectiveness of mathematics in this episode qualifies as unreasonable, in spite of the magnificent character of the achievement (the general theory of relativity). First, the considerable development of differential geometry in the 19th century shows no signs of a pre-established link between mathematics and physical needs. Rather, this development matches what we have stressed before: the freedom of mathematics from physical content and the concomitant possibility of

evolution free from external influences. Indeed, Gauss's theory fits in perfectly with the historical tradition of work in geometry. The abstract character of mathematics made it subsequently possible and natural to construct a geometrical theory of spaces of an arbitrary number of dimensions, *in spite of* the fact that this notion seemed completely superfluous in physics.

Second, differential geometry did not inspire relativity in its initial stages. It is true, as demonstrated by Minkowski, that already special relativity can be regarded as a geometrical theory of a four-dimensional space-time manifold. But mathematics did not really anticipate this application of its concepts to physics (a point regretfully noted by Minkowski in his essay). The geometrical approach did not play a role in the genesis of special relativity, and it took Einstein considerable time to recognize the value of the geometrical viewpoint. Indeed, one can very well defend the viewpoint that Einstein's original three-dimensional treatment is closer to physical experience than the abstract four-dimensional approach. The situation is similar to the ones discussed above: there are more ways than one to formulate special relativity mathematically, and it cannot be decided beforehand which way will proffer the best chances of fruitful generalization. But one *can* see beforehand that once the geometrical formulation is taken seriously, going from flat Minkowski space-time to curved Riemannian space-time constitutes a way of generalizing special relativity; this generalization is in its mathematical essence identical to what Gauss did in going from flat to curved surfaces. So differential geometry is evidently a suitable instrument to achieve one type of generalization of special relativity.

Summing up, the development of differential geometry can be understood from the internal dynamics of mathematics, without reference to its later application in relativity. The mathematics of differential geometry did not play a role in the genesis of special relativity. After the special theory had been developed, it turned out that differential geometry could be used as a tool—but that mathematics is able to give a geometrical description of special relativity cannot be considered remarkable, given its nature of a theory of invariants. It was not obvious beforehand that the use of differential geometry, and the type of generalization of special relativity suggested by it, would lead to a revolutionary new physical theory. Indeed, many physicists made attempts to incorporate gravitation into relativity in a non-geometrical way. That the application of differential geometry to relativity was in fact highly successful is very understandable with hindsight, given that general relativity has made it clear that the space-time of our world can be treated as a curved manifold. But the latter is almost tautological, and does not point into the direction of a pre-established

harmony between the developments of mathematics and physics. If one of the other research programs that were pursued after 1905 (e.g., Abraham's or Lorentz's) had been successful, the geometrical approach might have been forgotten by now.

One might answer that it is still an unreasonable coincidence that the geometrical tools lay ready just in time, waiting for Einstein to come along. I do not think that this is a convincing manoeuvre, however. We already saw that it is of the essence in mathematics that developments take place freely, in diverse directions. There is a steady rate of addition of new tools to the mathematical repertoire. In line with this, cases in which there are several mathematical approaches to choose from abound in the history of physics, and are not surprising. The genesis of relativity theory exemplifies this situation: traditional mechanical techniques competed with the geometrical approach. However, situations in which no suitable mathematical tools are available certainly occur too. For example, in present-day elementary particle physics physicists feel obliged to develop their own specialized new mathematics, adapted to the particular needs of string and membrane theories. This underscores the fact that the development of mathematics is not tuned to needs about to arise in physics.

5. TRANSPORTING INSIGHT

It is a well-known phenomenon that mathematical models and techniques used in the context of one physical problem are often also applicable to completely different areas in physics. This is made possible by the neutrality, in the sense of freedom of physical content, of mathematics: the same mathematical objects and symbols can receive completely different physical interpretations. Results achieved in one context can thus be translated to other contexts. For example, the same equations apply to electrostatics and laminar flow in fluids; these very different phenomena can both be regarded as models (in the sense of model theory) of the equations.

Mathematical correspondences between different fields are often used for illustrative purposes, for instance in physics education. But, importantly, they also play a significant role at the forefront of physical research, in breaking new ground. An interesting example from recent research in the foundations of physics is provided by a translation of the famous Bell theorem to a space-time context.

Bell's theorem demonstrates that the measurement results that are predicted by quantum mechanics cannot be interpreted as simply mirroring system properties that already existed independently of the measurements. This is to be contrasted with the situation in classical physics. According to

classical mechanics, particles possess properties like position and momentum quite independently of whether any measurements of these quantities take place. *If* measurements are made, the results should of course reflect the values that were already there. But as just said, this cannot be maintained in quantum theory. Here, the outcome of a measurement is in general not the reflection of an object system property that was already there; and what is more, it cannot even be considered to be independent of measurements performed far away.

The latter point is illustrated by experiments of the Einstein-Podolsky-Rosen type (Einstein, Rosen and Podolsky, 1935). In a modern version, two electrons whose total spin state is the singlet state, in which the total spin is zero, fly apart until their mutual distance has become very great. Subsequently, spin measurements are made on the individual particles. For each particle, there is the choice of measuring the spin in one of two directions. The experiment can be repeated with different choices of these directions, so that four combinations of directions will be measured in the series of repetitions. Correlations between outcomes in these four pairs of directions are predicted by quantum mechanics, and are verified in actual experiments. As is well known, these correlations violate the Bell inequality. Now, it is a mathematical fact that the Bell inequality is satisfied as soon as the spin values found in the measurements on the individual particles can be regarded as coming from one joint probability distribution of spin values (Fine, 1982a, 1982b). The latter would be the case if the individual electrons possessed spin values in all directions, as in classical theory, independently of which—or whether—measurements are going to be made. If that were true, there would be well-defined, definite spin values in the four directions under discussion in each run of the experiment; in repetitions of the experiment these values would vary and form an ensemble that defines a joint distribution of the four spin quantities. Only two of them could actually be measured in any single experimental run (one direction for each particle); the measured values would therefore be samples from this joint distribution. The violation of Bell's inequality by the predictions of quantum mechanics, and by the experimental results, shows that we cannot think of the EPR situation in such classical terms—the measurements do not reveal pre-existing jointly defined quantities.

In other words: the actually measured spin values cannot be considered to reveal local particle spins, independently of the kind of measurement performed on the other, far-away particle. See Figure 1 for a schematic representation of the situation: either σ_1 or σ'_1 is measured on electron 1, and σ_2 or σ'_2 on electron 2. The two horizontal and two diagonal lines symbolize the four possible combinations of measurements. The vertical double lines represent the electrons.

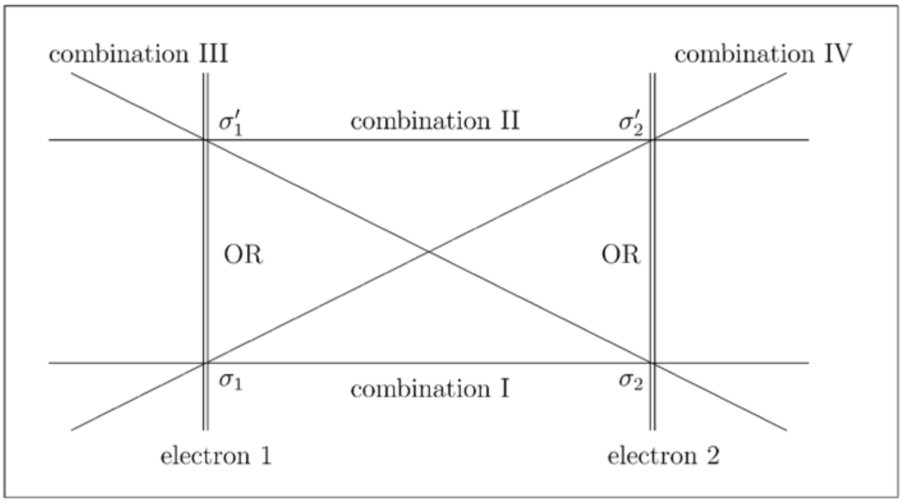


Figure 1. The four possible combinations of spin measurements.

This result is shocking for the classical intuition. It undermines the classical concept of locality, and even the very concepts of physical properties and physical systems; its ramifications have not been completely digested yet. But it turns out that there is even more in store. Exactly the same mathematical structure can be recognized in a new situation, so that the argument can be repeated there—with results that appear even farther-reaching (Myrvold, 2002).

Consider two well-localized systems, S_i, $i = 1,2$. Let α and β be two hyperplanes of simultaneity for some reference frame Σ. Let E_i be the places where the systems S_i are located on α, and let F_i be the corresponding regions on β (see Figure 2). We assume that the two systems are sufficiently far apart that E_1 is spacelike separated from F_2, and E_2 is spacelike separated from F_1. Let γ be a spacelike hypersurface containing F_1 and E_2, and let δ be a spacelike hypersurface containing E_1 and F_2.

If S_1 and S_2 are isolated during their evolution between α and β there will be unitary operators U_i such that the state of the combined system $S_1 \oplus S_2$ on β will be related to its state on α by

$$\rho(\beta) = U_1 \otimes U_2 \, \rho(\alpha) \, U_1^\dagger \otimes U_2^\dagger \tag{1}$$

If the regions E_1, E_2, F_1, F_2 are sufficiently small, they may be treated as points, and we may regard γ and δ as hyperplanes of simultaneity for reference frames Σ', Σ'', respectively.

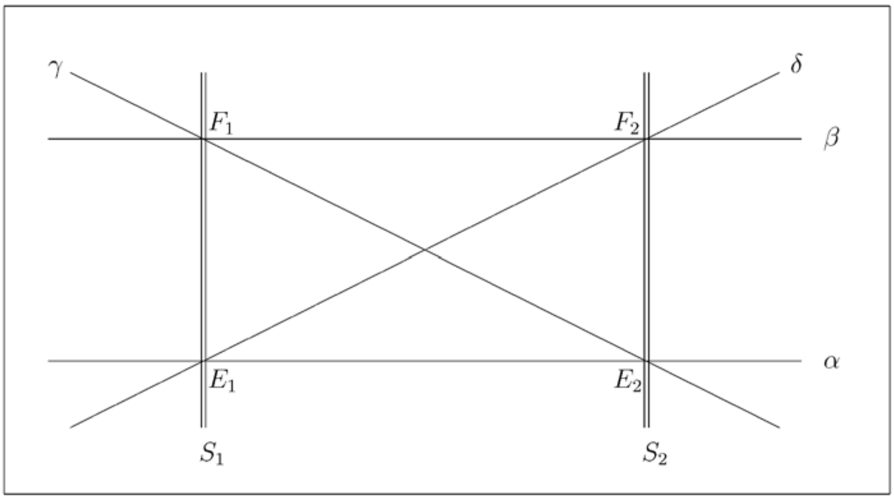


Figure 2. The four simultaneity hyperplanes α, β, γ and δ.

Let $\rho(\gamma)$ be the state on hypersurface γ, and let $\rho(\delta)$ be the state according on δ. On the basis of the assumption of unitary evolution between α and β, the states on the hyperplanes γ and δ can easily be related to $\rho(\alpha)$. We find:

$$\rho(\gamma) = U_1 \otimes I_2\, \rho(\alpha)\, U_1^{\dagger} \otimes I_2 \tag{2}$$

and similarly

$$\rho(\delta) = I_1 \otimes U_2\, \rho(\alpha)\, I_1 \otimes U_2^{\dagger}. \tag{3}$$

Now suppose that A_1 and A_2 are definite properties of S_1 and S_2, respectively, on α, and B_1 and B_2 are definite properties on β. This supposition fits in with interpretations of quantum mechanics according to which the quantum state assigns probabilities to objectively existing quantities (Bohm's interpretation or modal interpretations, for instance (Bub, 1997; Dieks and Vermaas, 1998)). Suppose further that the value of A_1 possessed by S_1 at E_1 is possessed by it without reference to the hypersurface containing E_1 that is contemplated, and similarly for the other points of intersection E_2, F_1, F_2; this is just the almost self-evident assumption that what happens at these four points are objective events located in space-time. There must then be a joint probability distribution over the values of our four observables, that yields as marginals the quantum mechanical Born probabilities on all four hyperplanes. In this we have

assumed the central tenet of special relativity, namely that the different frames of reference are equivalent; in our case that the Born probability rule applies equally on α, β, γ and δ.

But the states on the various hyperplanes are interrelated, as indicated in Eqs. (2)-(3). By inspection of these relations we find that the existence of such a joint distribution is equivalent to the existence of a joint distribution calculated in one state, namely $\rho(\alpha)$, and yielding, as marginals, the statistics for the observables $A_1 \otimes A_2, A_1 \otimes C_2, C_1 \otimes A_2, C_1 \otimes C_2$ where

$$C_i = U_i^\dagger B_i U_i \,. \tag{4}$$

However, as we have explained for the case of the EPR-experiment, such a joint distribution of four non-commuting observables, yielding the quantum mechanical Born marginals for the pairs of observables, cannot exist in general (Fine, 1982a; Fine, 1982b). Bell inequalities can be violated if there are no restrictions on the state, and the violation of a Bell inequality entails the non-existence of a joint distribution. Therefore, if $\rho(\alpha)$ is a state such that a Bell inequality can be derived for the observables A_1, C_1, A_2, C_2, then it cannot be the case that A_1 is objective at E_1, A_2 is objective at E_2, B_1 is objective at F_1, and B_2 is objective at F_2.

The argument here completely mimics the earlier Bell argument: the mathematics is identical. The structural isomorphy of the two arguments can clearly be seen from the similarity between Figure 1 and Figure 2. The symbols have different meanings, but the mutual relations are the same. Whereas in the original Bell case locality was at issue, we now find that it must make a difference whether we consider what happens in E_1, e.g., from the perspective of E_2 or from the perspective of F_2. In other words, events are not just there, but are different depending on the hyperplane of which they are considered a part. This result is a lot more perplexing than the original Bell non-locality conclusion! In the Bell case a property could be considered to depend on what kind of far-away measurement was made. But since only one such measurement can actually be made, no conflict arises with the uniqueness and objectivity of the property in question. In our new case, however, all the different contexts, i.e. the different hyperplanes, are jointly present. So, *events cannot in general be unique and objective in themselves* according to this quantum mechanical scheme, but must depend on the hyperplane on which they are considered to lie: a truly amazing conception.

6. CONCLUSION

> The previous section illustrates how mathematical arguments and techniques that are elementary and well-known in themselves can lead to unexpected new results, new ideas, and new directions of research. In the concrete case discussed, indications are that the very concept of an 'event' has to be modified in quantum mechanics. Quite generally, it appears that properties of physical systems are relational in character—that they are perspective-dependent (Bene and Dieks, 2002). This conclusion is tentative, and further research concerning these issues is needed. It is not the purpose of this article to argue for specific theses concerning the nature of quantum mechanical reality.

Rather, what I wanted to show is how mathematics, by its very nature of a subject without physical content, lends itself to an unlimited variety of applications. As soon as some type of order, structure or regularity is present in an area, mathematics becomes almost automatically useful. Because of the strong internal dynamics of the discipline, and the steady growth of its repertoire, it is not unlikely that some suitable mathematical technique is already available when new fields of physical research are opened up. If not, this will be an impetus to develop new mathematical tools for the purpose at hand.

It would be wrong to summarize this by saying that mathematics is *nothing but* a descriptive instrument, which can be employed in many circumstances. Because mathematics is so expressive and rich in conceptual tools, it transcends the role of just a language; it can sometimes actively lead the way in physical research. More accurately, mathematics does not impose directives about how to proceed; rather, the abundance of mathematical instruments makes it possible for researchers of all sorts and inclinations to proceed along the ways of their own liking. In this process, mathematics can suggest generalizations and new directions, as illustrated by the cases of holism and relativity. It can also provide new insight by transporting old results to new contexts. It can thus make for 'miraculous' progress, even though its effectiveness is no wonder at all.

REFERENCES

Bene, G., and Dieks, D., 2002, A perspectival version of the modal interpretation of quantum mechanics and the origin of macroscopic behavior, *Foundations of Physics* **22**: 645-671.

Bub, J., 1997, *Interpreting the Quantum World*, Cambridge University Press, Cambridge.

Dieks, D., and Vermaas, P. E., 1998, *The Modal Interpretation of Quantum Mechanics*, Kluwer Academic Publishers, Dordrecht.

Einstein, A., Rosen, N., and Podolsky, B., 1935, Can quantum-mechanical description of physical reality be considered complete?, *Physical Review* **37**: 777-780.

Fine, A., 1982a, Hidden variables, joint probability and the Bell inequalities, *Physical Review Letters* **48**: 291-295.

Fine, A., 1982b, Joint distributions, quantum correlations and commuting observables, *Journal of Mathematical Physics* **23**: 1306-1310.

Myrvold, W., 2002, Modal interpretations and relativity, *Foundations of Physics* **32**: 1773-1784.

J. von Neumann, J., 1932, *Mathematische Grundlagen der Quantumtheorie*, Springer, Berlin.

Russell, B., 1917, Mathematics and the Metaphysicians, in *Mysticism and Logic*, George Allen and Unwin, London.

Wigner, E., 1960, The unreasonable effectiveness of mathematics in the natural sciences, *Communications in Pure and Applied Mathematics* **13**: 1-14.

Wu, T. T., and Yang, C. N., 1975, Concept of nonintegrable factors and global formulation of gauge fields, *Physical Review D* **12**: 3845-3857.

THE LAWS OF NATURE AND THE EFFECTIVENESS OF MATHEMATICS

MAURO DORATO
University of Rome 3, Rome, Italy

Abstract: In this paper I try to evaluate what I regard as the main attempts at explaining the effectiveness of mathematics in the natural sciences, namely (1) Antinaturalism, (2) Kantism, (3) Semanticism, (4) Algorithmic Complexity Theory. The first position has been defended by Mark Steiner, who claims that the "user friendliness" of nature for the applied mathematician is the best argument against a naturalistic explanation of the origin of the universe. The second is naturalistic and mixes the Kantian tradition with evolutionary studies about our innate mathematical abilities. The third turns to the Fregean tradition and considers mathematics a particular kind of language, thus treating the effectiveness of mathematics as a particular instance of the effectiveness of natural languages. The fourth hypothesis, building on formal results by Kolmogorov, Solomonov and Chaitin, claims that mathematics is so useful in describing the natural world because it is the science of the abbreviation of sequences, and mathematically formulated laws of nature enable us to compress the information contained in the sequence of numbers in which we code our observations. In this tradition, laws are equivalent to the shortest algorithms capable of generating the lists of zeros and ones representing the empirical data. Along the way, I present and reject the "deflationary explanation", which claims that in wondering about the applicability of so many mathematical structures to nature, we tend to forget the many cases in which no application is possible.

Key words: mathematics; laws of nature; algorithmic complexity theory; evolution; semantics.

> Our experience hitherto justifies us in believing that nature is the realization of the simplest conceivable mathematical ideas.
> (Einstein,1933)

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 131–144.

1. INTRODUCTION

In this note I will try to connect the difficult question of the effectiveness of mathematics in the natural science (Wigner, 1967) with the philosophical issue of the nature of natural laws. Trying to create a bridge between these two as-yet unrelated areas of philosophical research seems a fruitful enterprise as it could shed light on both. On the one hand, for example, the philosophical questions of (i) what laws *are* (ontological and semantic realism *vs.* ontological and semantic antirealism about laws) and (ii) how we come to know them, seem to be questions that can be fruitfully approached anew if we pay attention to the mathematical character of laws: the idealized, abstract and simplified character of laws in our models is often due to the need of having *tractable* mathematical problems and solutions. This is, for instance, why *stable* solutions to *linear* differential equations were privileged at the beginning of modern mathematical physics. On the other hand, the "unreasonable effectiveness of mathematics" is an abbreviated slogan to refer to the *descriptive, predictive and explanatory power of mathematics in dealing with the natural (and social) world*, a power that almost exclusively depends on the fact that laws in most natural and social sciences are expressed in mathematical language. The fact that scientific descriptions, predictions and explanations are often a matter of *subsuming single phenomena under laws* (together, of course, with a specification of boundary conditions and/or initial conditions), confirms the importance of establishing a relationship between the problem of the effectiveness of mathematics and the philosophy of the laws of nature.

As a further illustration of the opportunity of a cross-fertilization of the two areas above, consider that the ontological realist about laws believes that there are mind-independent truth-makers (for instance, relations about properties of physical systems) making any well-confirmed law-statement true. What is the character of such truth-makers? It is an interesting and not-yet sufficiently explored question whether the entities referred to by such statements are describable in a sufficiently faithful way in a non-mathematical language. If this were not feasible, we might have to conclude in a Pythagorean fashion that the ultimate structure of the universe is mathematical in character. As Steiner put it (Steiner, 1998):

> Can we specify, by using a non-mathematical language, how can the world be made in such a way that valid mathematical deductions are effective in predicting observations? (p. 24)

I should add that rather than offering alleged "definitive solutions" to the hard problems raised above, the paper tries to suggest new directions of inquiry by reviewing and briefly evaluating the scanty available literature. In

the next section, I will try to provide arguments to the effect that the problem of the effectiveness of mathematics is a *genuine* one. Such a preliminary task is necessary, since if the problem posed by Wigner were a pseudo-problem, the claim that it could help us to look at the issue of scientific laws under a new light would be groundless. In the third section, I will present and briefly evaluate the current attempts at answering Wigner's problem (I will list four of them). In the fourth section, I will focus on one of these attempts, centred on the view that laws of nature are the *software* of the physical universe (algorithmic view of the laws of nature and of the applicability of mathematics). In this metaphor, which for its proponents is suggestive of a deeper truth about mathematically formulated laws, the universe is considered to be a gigantic computer whose *hardware* is whatever fundamental physics tells us about the ultimate component of matter (fields, particles, superstrings, etc) and whose software abbreviates and compresses the ordered sequence of states it goes through in time. In the fifth section, I will raise various difficulties to the algorithmic conception of laws, some of which may appear fatal to the whole project.

2. SOME SCEPTICAL REMARKS AGAINST WIGNER'S QUESTION

As is often the case in philosophy, one often wonders whether an apparently deep problem to a closer analysis might reveal itself but a pseudo problem. In our case, a dissolution of the question why mathematics can be used to describe, predict and even explain physical phenomena – think of the structural-geometrical *explanation* of the time-dilation or of the length-contraction effects in terms of the *geometry* of Minkowski space-time, in which an invariant *four-dimensional* entity is projected onto different relative 3-spaces with its own inertial time – is the remark that *the effectiveness of mathematics meant in this sense results from a selective effect*, *making us focus only on the evidence of success*. Ignoring evidence of failure, one can make any hypothesis look good, so that one could as well refer to the un-effectiveness (or the very frequent failure) of many mathematical theories to apply to the natural world. A second objection consists in the fact that very often, as in the case of calculus at the times of Newton and Leibniz, pieces of mathematics are invented and constructed with the intention to solve physical problems: no wonder that they sometimes work!

In order to tackle the first objection, note that we cannot take statistics in order to determine the fraction of areas of mathematics that have had a successful application over the total number of areas of mathematical

research. Therefore, since the objection presupposes such a statistical count, it loses its force. However, even if it did make sense to take such a counting, we would not find many branches of pure, non-applied mathematics that are not somehow connected with the empirical world.[49]

Of course, we must agree that we can apply mathematics to the empirical reality only in some few, lucky cases and not in others, namely when we are dealing with simplified or "simple" enough phenomena[50]. Cases in which the mathematician cannot find the way to describe a natural process clearly outnumber those in which she can (see Steiner, 1998, p. 9).

However, I take it that it is still mysterious that *some* of the consequences of our mathematical symbols are also consequences of the symbolized phenomena, as Hertz once put it, especially when the symbols were not created for applicative purposes. If the application of a part of mathematics to physics has been, at least sometimes, unpredictable and unexpected – that is, if that part of mathematics had been constructed for pure, or non-applicative purposes – *then I claim that even a single instance of successful applicability would cry for an explanation*.

This remark takes us to the second reply to our question: for instance there are parts of mathematics like group theory that had not been invented to deal with physical problems, but nowadays it would be difficult to deal with the zoo of elementary particles without using the algebra of groups. The use of Levi Civita's and Ricci's absolute differential calculus in the general theory of relativity is an often quoted instance of a surprising spin off that had not been pursued intentionally. Dirac discovered his equation (and the existence of antimatter) by working on a formal, syntactical analogy with what Pauli had done before him with 2x2 matrices: despite the fact that Dirac was trying to solve a physical problem, it is still mysterious that merely formal analogies like these can sometimes be conducive to truth or success. In his characteristically direct and non-pompous style, Feynman put it thus (Feynman, 1967):

> I find it quite amazing that it is possible to predict what will happen by mathematics, which is simply following rules which really have nothing to do with the original thing. (p. 171)

Finally, we should also note that an explanation for the applicative success of mathematics cannot be easily found just by taking a philosophical stance on the ontology of mathematical objects. In fact, just to name two

[49] By "connected" I mean to refer also to cases like the application of number theory to the problem of secret coding.
[50] Simple here is almost a synonymous of "can be captured by a mathematical model".

positions, the problem of the applicability of mathematics creates troubles for both the constructivists and the platonists. The former must explain why a *creation* of ours, often pursued for subjective purposes and without any pragmatist interest, can carry so much descriptive and predictive power, enabling us to explain and classify entities of the natural world that we did *not* create (under the typical assumption of the scientific realist). The platonists should explain why the natural world should conform or be partially isomorphic to the abstract realm populated by mathematical facts and entities, besides answering the usual objection coming from the causal theory of knowledge. If the abstract real is causally inert, how can we get to know it?

3. FOUR POSITIONS ON WIGNER'S QUESTION

3.1 Steiner's antinaturalism

The first position I will list, which is *antinaturalistic*, has been explicitly defended by Steiner (1998): the universe is "user friendly" for the mathematician, but no explanation in natural terms seems available. Here, anthropomorphism strikes back, with a theological undertone (Steiner, 1998, p.10). According to Steiner (1998):

> to use mathematics to define similarity and analogy in physics is almost as anthropocentric as using "male-female" or "earthly-heavenly" as classifying tools. Why? Because the concept of mathematics itself is species-specific. There is no objective criterion for a structure to be mathematics – and not every structure count as a mathematical structure…Mathematicians today have adopted internal criteria to decide whether to study a structure as mathematical. Two of these are *beauty* and *convenience*. (pp. 8-9).

The interest of Steiner's book lies more in providing interesting historical examples of unexpected applications of mathematical concepts to the physical world via mathematical, formal analogies than in trying to explain the problem at stake. Concluding that the effectiveness of mathematics amounts to an anthropomorphic ("childish-like") projection of our constructed structures onto the physical world is simply a re-description of what needs to be explained, namely that mathematics, viewed as a social construction, somehow miraculously captures parts of a universe that is independent of our minds.

Another "risk" of Steiner's antinaturalistic position is that of encouraging a sort of mysticism about Wigner's problem: if naturalism means (i) that the existing or the real coincides with what exists in space and time, (ii) that we have the power in principle to investigate and get to know reality *via* the discovery of natural laws, (iii) that we can also find out how we get to know the world, antinaturalism might mean that we will never find out through science why our mathematically formulated laws enable us to predict and classify natural entities. This would somehow mean the bankruptcy of the possibility of coming up with an explanation to our question. We can therefore move on to a different attempt at explaining the problem.

3.2 The Kantian answer

The second position could be called a form of "kantism": mathematics works because, roughly speaking, we perceive the world by using inner mathematical intuitions (space and time), that is, we project *a priori* forms, constructions and categories of our own unto our experience of nature. Here the explanation of the effectiveness of mathematics can rely on data coming from evolutionary and cognitive psychology and as such it is not at all antinaturalistic (Longo, 2000a, 2000b). On the contrary, its promising aspect is the fact that it can be measured against in principle available *empirical evidence*, like data on arithmetical abilities of other animals, based on the evolutionary advantage of being able to distinguish, for example, "one predator" from "many predators" or even "one" from "two" or "three" (Devlin, 1999)

Just to offer an argument based on mere *a priori* plausibility of a Kant-type explanation, one could reason as follows. Since *mathematical abilities* are to a good extent genetically determined, fundamental mathematical concepts, like *number* or *space*, might be *a priori*, in the same sense in which fundamental concepts are a priori in Fodor's *language of thought*. Otherwise, one could ask, what would be the *object* of such mathematical abilities? Furthermore, *if* the contents of our thoughts are expressed in symbolic structures of an innate language, whose syntax and semantics are similar to (though more abstract than) those of the natural language, *then* the claim that all our mental processes are essentially *computational* could explain why the development and justification – not thought necessarily the origin of certain concepts – of mathematical knowledge is *a priori* without invoking any outlandish form of Platonism.

Note that the question of applicability within this second attempt would be solved only in part, because while it can be true that we perceive the world also thanks to "primitive mathematical intuitions", it remains to be

explained why developments of mathematics that are very remote from those intuitions can still have an empirical application. The question of the nature of laws of nature is therefore not even touched by this second attempt.

One should in fact consider that mathematics is applicable also at scales (like those typical of atoms and subatomic particles on the one hand and cluster of galaxies on the other) that differ by many orders of magnitude from the dimension of the physical bodies to which we adapted during our biological evolution. The difficulty then is as follows: if one believed in this sort of Kantian explanation, supplemented by whatever part of evolutionary psychology or cognitive psychology may come to help, one would have to say that it is only through analogy that we can extend what works at our dimension to other, much smaller or much bigger dimensions. But notoriously the analogies between the laws of the atom and those of classical physics break down, and we certainly have no direct experience of atoms and molecules. And yet mathematics applies successfully also to the quantum world, which in Kantian terms is a sort of noumenal world going beyond experience. It seems that the Kantian position needs to answer this objection, at least to the extent that it is plausible to assume, as it is, the mind-independent existence of atoms and molecules:[51] if mathematics applies to the objects of our experience because the latter is possible *only via* the primitive mathematical notions that belong to our subjective side, how can we apply mathematics so successfully to objects that are *beyond* the reach of experience? Why should evolution have equipped us with the laws of objects that, like atoms, play no role in our ordinary life?

A possible reply to this objection is suggested by Steiner: scientists extend the applicability of mathematics by using Pythagorean or syntactic-formal *analogies* between physical laws written in purely mathematical language, for which no translation into non-mathematical language is possible. This suggestion can have a two-fold interpretation. Either scientists correctly believe that the world is objectively written in mathematical language, as Galilei thought, or the success of mathematics in extra-phenomenical realms is utterly unexplainable, because it is the result of pure chance, the coincidence being given, for instance, by the fact that physical laws suggested by mathematical analogies with laws working within the domain of macroscopic bodies work also within the quantum domain.

In either hypothesis, however, the Kantian position is in trouble, because it makes the explanation of the effectiveness of mathematics impossible.

[51] The way out of denying the existence of mind-independent entities is of no use, since the predictive power of structures like that of the Hilbert space is undeniable and needs to be explained.

3.3 Mathematics as a kind of language

The third position is very close to the second, and argues that mathematics is a type of language, so that the question of the applicability of mathematics is a chapter of a generalized theory of semantics. Ordinary language is successful in describing, predicting and explaining many properties of ordinary objects: mathematics is just a sophistication of these abilities, since relative to the natural languages, mathematics is just a more rigorous, less ambiguous, and formally organized language. Chomsky's generative linguistics and Fodor's computational view of human thought (Fodor, 1975) may give support to this third position, since besides the algorithmic nature, mathematics seems to share with both language and thought the characters of *productivity and systematicity*. The number of mathematical results and theorems that can be generated from certain premises seems potentially infinite, and mathematics keeps growing in many different directions simply by building in a recursive and combinatorial way on previous results (productivity). All branches of mathematics are deeply connected with each other, so that the capacity to generate a certain result is intrinsically connected to the capacity of generating other results (systematicity).

However, a linguistic approach to the problem of effectiveness of mathematics is affected by the difficulty of finding non-mathematical correlates for central mathematical concepts. This is certainly *not* the case with ordinary languages, whose referential terms have always a well-defined extension in the outer world. Consider the following list of notions, characterized by an increasing abstraction. We do not have too many troubles in finding a non-mathematical correlate for the mathematical notion of *subtracting* two numbers, since it corresponds to the physical operation of separating objects that were previously together; likewise, the mathematical notion of "*linearity*" corresponds to the superposition of two waves or to the fact that two causes contribute separately to their effect. *Fiber bundles* may be taken to describe gauge fields, but what does the *analyticity* of a function correspond to in the real world?[52] While the latter mathematical notion is crucial in the applicability of many parts of mathematics, it does not seem to have any counterpart in the physical world. Once again, unless we assume that any important mathematics notion can be parsed in a non-mathematical language, which seems difficult, we have no way of making sense of the applicability of the symbolic language of mathematics to the physical world. A literary quotation addressing the role of complex numbers, due to the

[52] Some of these examples are due to Steiner (1998).

German writer Robert Musil, will conclude this brief discussion of the third position (Musil, 1906):

> The strange fact is that with these imaginary or even impossible numbers one can anyway make perfectly real calculations which end in a concrete result. (p. 56)

Ironically, at the time of *The Confusions of the Young Törless*, from which this passage is taken, Musil could not be aware at the fact that the most successful theory of the atomic structure of matter – quantum mechanics – would have been using complex, "imaginary" numbers to calculate the probability of measurements.

3.4 Mathematics as the science of the abbreviation of sequences

We can regard a mathematically formulated law as a "bridge" colligating two banks of a river, each constituted by quantitative data resulting from measurements. On one side of the river we find the initial data or boundary conditions, which in our metaphor we can regard as the *input*, and on the other side we find the predictions or retrodictions – the *output* – the result of a *calculus*.

Since such a result is obtained in a purely deductive fashion, that is, thanks to the application to the initial data of a mechanical rule given by the physical law, the metaphor of the scientific laws regarded as the algorithm of a computer appears initially justified. If the initial data in fact are such as to satisfy some mathematical conditions which in this context can be omitted,[53] and whenever the solution to the equation exists and is unique, a mathematical law expressed as a differential equation enables us to transform in a finite numbers of steps, and in purely mechanical fashion, the initial data in final predictions or retrodictions (output).

What interests us is, of course, whether such an analogy between the laws of succession of any physical system – regarded as something that evolves in time by going through a finite number of states describable in physical language – and the software of a computer, can help us to better understand: (i) why the world is describable by mathematical laws and (ii) how the latter are related to the world, that is, how they *represent* it. In order to shed light on the presuppositions of the law-software analogy, we should ask whether also a physical system, in a sense to be specified, could be said

[53] The functions representing the data must be differentiable at least as many times as the degree of the differential equation giving the algorithm.

to "compute" its "next" state *by bringing it about*. Let us assume, for the sake of the argument, that it makes literal sense to claim that a physical system going from an initial to a final state *literally executes a program or calculates its future state*, in the same sense in which a mathematical physicist deduces or calculates the predictions corresponding to the initial data.

We will now discuss more closely this view of the applicability of mathematics, in order to evaluate its chance to be a candidate to meet Wigner's challenge.

4. THE ALGORITHMIC VIEW OF THE LAWS OF NATURE

Suppose that the temporal evolution of any physical system is describable by *finite strings of real numbers*, corresponding to operational measures of physical magnitudes (temperature, pressure etc.). We can have two cases: such strings can be ordered

(111000111000111000…)

or truly random

(0100110101100110…).

In the first case, the string can be generated by a simple instruction ("print 111000 n times"), which is much shorter than the list itself. In the second case, the string *appears* as truly random, where "appear" is meant to stress that while we can show that a finite string is not random by giving the generating law (algorithm), we can never prove that a string is random (this is a version of the halting theorem).

At this point we can give two definitions, based on algorithmic complexity theory, which will be relevant for our purpose:

Definition1: the complexity of a string is the length of the shortest algorithm capable of generating it

Definition2: a string is said to be algorithmically compressible when there is an algorithm capable of generating it, such that its information content (number of bits) is much less than that of the string.

As an illustration of these definitions, consider that a string like

$$\{1, 4, 9, 16, 25, 36, \ldots\} \tag{1}$$

is obviously not random, since it can be trivially obtained by squaring the positive integers in the list

$\{1, 2, 3, 4, 5, 6, \ldots\}$ (2)

If the numbers in (2) correspond to measured magnitudes in such a way that, say in a temporal interval of 1, 2, 3 seconds (the input data), a body travels 1, 4, 9 meters (the output), then the existence of a rule generating (1) from (2) shows that (1) *is algorithmically compressible*. The above algorithm is – *modulo* the constant $\frac{1}{2}g$. – Galileo's law of free fall, generating the spatial intervals (1) from the square of the temporal intervals (2).

In a word, by following the metaphor of scientific laws regarded as the software of a physical system, we discover that *searching for laws is tantamount to asking which is the length of the shortest program capable of generating the string of numbers expressing the experimental measures*. Such a length – the complexity of the string – will be equal to that of the original string only if the latter is composed by apparently random numbers, and does not obey any known law.

The idea that scientific laws are an economic synthesis of all the information contained in our observations is certainly not new, and in this algorithmic approach it finds a new, rigorous and precise formulation. It was especially Ernst Mach who regarded science and its theories and laws as an economic "summary" of our observations. As he wrote (Mach, 1896):

> Science is a form of business. Its purpose is to find the maximum amount of the infinite eternal truth with the minimum amount of work, in the minimum expenditure of time and with the minimum amount of thought effort. (p. 14)

After having made explicit the philosophical consequences that seem to follow from the software metaphor for scientific laws, we can now finally discuss a possible explanation of the applicability of mathematics, due to the physicist John Barrow (Barrow, 1992):

> science exists because the natural world seems algorithmically compressible. The mathematical formulae that we call laws of nature are economical reductions of enormous sequences of data expressing changes of state of the world: here is what we mean by intelligibility of the world…Since *the physical world is algorithmically compressible*, mathematics is useful to describe it because it is the language of the abbreviation of sequences. The human mind enables us to make contact with that world because our brain has the ability of compressing complex sequences of sense data in shorter form. Such abbreviations make thought and memory possible. The natural limits that nature poses to our

senses prevent us from overloading our brains with information about the world. Such limits are security gates for our minds. (p. 93-96)

5. SOME DIFFICULTIES OF THE ALGORITHMIC VIEW

In discussing some of the difficulties generated by this position, let us start to discuss parts of Barrow's quotation.

(1) On the "epistemic" hand, if it is the brain that filters sense data and elaborates them through the construction of mathematical concepts, what is the relationship between such a capacity and the applicability of mathematics, regarded as the technique of compressing sequences? This is an open question and it connects this approach with the second (and third) position. Until we have a clear answer to this question, it is difficult to credit Barrow's claim for being more than an interesting speculation.

(2) On the "ontic" hand, it is not clear what it means to affirm that "the physical world is compressible": isn't this another way of formulating what we are trying to explain? Furthermore, note that compressibility is an *epistemic* notion: we are interested in compressing information, nature isn't. However, we are after an *ontological* interpretation of laws regarded as algorithms, enabling us to understand why mathematics is applicable with success.

(3) As persuasively shown by McAllister in a different context (McAllister, 2003), strings of data have interest-relative patterns. In his example, McAllister asks us to consider a very long string of data on the atmospheric temperature: we can find many regular patterns with different periods. We will find regularities with a period of a day (due to the rotation of the earth), other patterns lasting some days (due to the weather systems), one lasting a year (due to the earth's revolution around the sun), patterns which are repeated every 11 years (due to the sun spots), other regularities which are 21000 years long (due to the precession of the earth's orbit). Each of these patterns has a different algorithmic complexity. Which is the *intrinsic* algorithmic or effective complexity of the string of data? Since the answer to this question is interest-relative, this approach cannot be used for any ontological purpose.

(4) Not all laws of nature are sequential (i.e., laws of succession) as the notion of laws as algorithm requires! Since the notion of algorithm is essential temporal (even in parallel-distributed computation, the results of the distinct calculations must interact before the output), either all natural laws are laws of succession, or else natural laws cannot be given an

ontological interpretation within the algorithmic view, because *ontologically interpreted laws have to be executed by physical systems*. Tale laws such as

$$F = G(M_1\ M_2)/r^2 \tag{3}$$

$$\Phi = e_0\ q \tag{4}$$

$$PV = kT \tag{5}$$

Either they all derive from laws of succession, or physical systems cannot be said to instantiate them, because their parts would have to communicate at superluminal speed. A law of coexistence in fact links by definition parts of physical systems that are spacelike-related.

Before concluding, it is important to clarify what this objection does *not* entail. Clearly, we *can* use the laws in Eqs.(3)-(5) to calculate one side of the equation from the other, and in this sense their algorithmic character is obviously not refuted. However, calculations to humans take time, and note that – in order to give the current interpretation of laws an ontic significance – we were assuming that physical systems do go from one of their states to another one by executing an algorithm. It is in this sense that only temporally related states of a system can execute an algorithm and only laws of succession can be captured by the algorithmic view of the laws of nature.

6. CONCLUSION

As it should be obvious from the above survey, the problem of the effectiveness of mathematics is here to stay, and no one of the solutions that we have sketched here is devoid of serious difficulty. Clearly, research in this stimulating area calls for a multidisciplinary effort, coming from philosophers of mathematics, historians of philosophy, epistemologists, linguists, historians of science, cognitive scientists and possibly neuro-physiologists. And the problem is of immense interest also to physicists, as the contemporary Nobel prize Weinberg admits that “it is positively spooky that the physicist finds the mathematician has been there before him” (Weinberg , 1986, p. 725).

Finally, Wigner’s problem has all the trademark of a deep philosophical problem, since not only does it favour the dialogue between the science and

the humanities, but also helps us to understand the place of mankind in the universe.

REFERENCES

Barrow, J., 1992, *Perché il mondo é matematico*, Laterza, Roma-Bari.

Devlin, K., 1999, *The Math Gene***.** *How Mathematical Thinking Evolved & Why Numbers Are Like Gossip***,** W. H. Freeman and Company, New York.

Einstein, A., 1933/1995, On the method of theoretical physics, in *Ideas and Opinions*, Three Rivers Press, New York, pp. 273-274.

Feynman, R., 1967, *The Character of Physical Law*, The Mit Press, Cambridge Mass.

Fodor, J., 1975, *The Language of Thought*, Crowell, New York.

Longo, G., 2000a, The constructed objectivity of mathematics and the cognitive subject; http://www.dmi.ens.fr/users/longo.

Longo, G., 2000b, The reasonable effectiveness of mathematics and its cognitive roots; http://www.dmi.ens.fr/users/longo.

Mach, E., 1896/1986, *Popular Scientific Lectures*, Publisher Group West, Berkeley.

McAllister, J. W., 2003, Effective Complexity as a Measure of Information Content, *Philosophy of Science* **70**(2): pp. 302-307.

Musil, R., 1906/2001, *The Confusion of Young Törless*, transl. by S. Whiteside, Penguin Book, New York.

Steiner, M., 1998, *The Applicability of Mathematics as a Philosophical Question*, Harvard University Press, Cambridge Mass.

Weinberg, S., 1986, Lectures on the applicability of mathematics, *Notices of the American Mathematical Society* **33**(5): 716-733.

Wigner, E., 1967, The Unreasonable Effectiveness of Mathematics in the Natural Sciences, in *Symmetries and Reflections*, Indiana University Press, Indiana, pp. 222-237.

SOME MATHEMATICAL ASPECTS OF MODERN SCIENCE AND THEIR RELEVANT PHYSICAL IMPLICATIONS

The Subtle Interplay of Entanglement and Non-locality

GIANCARLO GHIRARDI

University of Trieste, The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy

Abstract: After having stressed the role of mathematics for the elaboration of one of the pillars of modern science, Quantum Mechanics, we point out one of its most striking features: entanglement. It stems from the fact that the Hilbert space appropriate for the description of such a system is the direct product of the Hilbert spaces of the constituents. The linear nature of the theory allows then the occurrence of states that are not the direct product of two states belonging to the factors of the total Hilbert space. Such states are entangled and their mathematical structure gives rise to extremely interesting quantum effects that exhibit nonlocal features. It is shown how the elaboration of the mathematical framework which is appropriate for the quantum world has led to consider entangled states and to discover their extremely peculiar and interesting features, the most important being actually nonlocality. We also prove that the nonlocal features of nature which are associated to entangled states do not derive from the specific formulation and interpretation of the theory but are unavoidable, just due to the tested correlations between far away states implied by the formalism. The subject represents an ideal arena to see how the mathematical formalism, when combined with the appropriate physical insight and with extremely refined experimental techniques, can lead to discover unsuspected and revolutionary aspects of natural phenomena.

Key words: Hilbert space formalism; entanglement; nonlocality.

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 145–160.

1. INTRODUCTION

I consider it extremely stimulating the subject of this year's Losinj meeting, i.e. to reconsider the role of mathematics in Physical Sciences, an expression to which I will attach a much more general meaning by reformulating it as “the role of mathematics for our understanding of the world around us”. In this spirit I will start by summarizing the peculiar intellectual experience of all those who have been seriously involved in the theoretical aspects of science by making reference to a relevant quotation by the great theoretician E. P. Wigner (1959) who decided to choose as the title of one of his most relevant lectures the following one:

The unreasonable effectiveness of mathematics in the natural sciences.

This statement expresses in an extremely significant and synthetic way the wonder of any researcher when he realizes how the formal and mathematical achievements of human thought represent an important key for penetrating the mystery of natural processes.

However, to make clear my position, I feel also the duty of reporting the opinion that Wigner himself has expressed in the closing sentences of the above paper:

> *The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. We should be grateful for it and hope that it will remain valid in future research and that it will extend, for better or for worse, to our pleasure, even though perhaps also to our bafflement, to wide branches of learning.* (p. 237)

I have reported the last sentence to specify that, even though I consider them perfectly legitimate, I do not share "Platonic" attitudes about our subject. In my opinion the fact that a specific mathematical formalism reveals itself as the appropriate and ideal language to account for **new** phenomena has, by itself, an extreme conceptual relevance and can lead to radical changes about our understanding of natural processes. On the contrary, I do not believe that looking for new abstract formalisms, i.e. doing,

“mathematics for mathematics’ sake”

can, by itself, open the way to revolutionary new intuitions about the world around us when it is not accompanied by a corresponding identification of new and unexpected physical phenomena. Unfortunately my friend Prof. A. Miller has not been able to participate to this conference. However, I cannot loose the opportunity of calling attention to his deep investigations on the processes of scientific discovery and in particular on the detailed analysis appearing in his book: *Imagery and Scientific Thought*

(Miller, 1987). Among many interesting topics, he gives a detailed account of the birth of our best theory, Quantum Mechanics. We can raise the question: what emerges from this extremely lucid analysis?

First, that some physicists, like N. Bohr with his love for "atomic orbits" or E. Schrödinger with his desire to stick to the illuminating suggestions of de Broglie and Einstein, have chosen

visualizability

as the key road to scientific knowledge, while others, like W. Heisenberg, have privileged

abstract and formal approaches

resorting to what Schrödinger himself has qualified as "transcendental algebra". However, Heisenberg did not discover "matrix mechanics" or better "linear infinite dimensional complete and separable Hilbert spaces" just by adventuring himself in formal speculations – actually this had been done years before by D. Hilbert within a precise mathematical line of research. Heisenberg's research program was firmly grounded on empirical data and he appropriately insisted on the need of disregarding the untestable statements about atomic orbits and/or positions (which unavoidably led to contradictions) claiming that one had to stick to the only clear empirical new facts contradicting the classical view: the quantization of physical quantities and the discrete nature of spectral lines.

Summarizing, I convincingly share the opinion of Galileo that:

> *Philosophy is written in that big book which is continuously open in front of our eyes (I mean the Universe), which however one cannot understand if, in advance, one does not master the language and one does not know the ciphers it uses. The language is mathematics and the ciphers are triangles, circles and other geometrical figures,*

but at the same time I consider extremely important to keep always in mind his fundamental motto:

Provando e riprovando

which invites us to be guided by experience before adventuring in wild speculations about *what is out there*.

To conclude these preliminary remarks I will therefore summarize my position as follows:

Experience, the basis of any scientific knowledge, suggests new and innovative theoretical and mathematical perspectives. When one such perspective reveals itself as particularly successful to account for some basic features of a revolutionary phenomenological framework one has to exploit all its subtle formal aspects since they can yield unexpected, new and innovative views about nature.

This situation is paradigmatically exemplified by quantum phenomena, the Hilbert space description of them, and in particular by its peculiar mathematical trait, entanglement, with its implications for the non-local character of natural phenomena. After this premises I can pass to the subject of my talk.

2. THE QUANTUM FORMALISM

The way quantum theory accounts for natural process is embodied in the following rules:

1. The states of a physical system are described by elements (statevectors) of a linear, complex, infinite dimensional complete and separable vector space, and, as such, they can be multiplied by arbitrary complex numbers and summed (*the superposition principle*).
2. The physical observables are represented by appropriate (self-adjoint) operators on this space. Any such operator $\hat{\Omega}$ identifies uniquely its spectral family, i.e., its eigenvalues ω_k – a subset of the real axis – and the associated eigenvectors $|\omega_k>$:

$$\hat{\Omega}|\omega_k>= \omega_k|\omega_k> \tag{2.1}$$

   The eigenvectors are a complete set, in the sense that any statevector $|\Phi>$ can be expressed as a linear combination (discrete and/or continuous) of them

$$|\Phi>= \sum_k c_k|\omega_k>+ \int_{cont.\,spectrum} c(\omega)|\omega>d\omega \tag{2.2}$$

   Assuming that $|\Phi>$ and the eigenvectors of $\hat{\Omega}$ are appropriately normalized, the coefficients of the development are given by the corresponding scalar products: $c_k =<\omega_k|\Phi>, c(\omega)=<\omega|\Phi>$, which satisfy $\sum_k|c_k|^2 + \int_{cont.spectrum} |c(\omega)|^2 d\omega = 1$.

3. The only possible outcomes of a measurement of the observable $\hat{\Omega}$ are the eigenvalues of the associated operator.
4. The preparation procedure of a system consists in measuring an observable and getting an outcome: after the measurement the statevector coincides with the eigenvector associated to the obtained outcome.

5. The evolution is governed by the (norm preserving) deterministic and linear Schrödinger's equation:

$$\hat{H}\left|\Psi,t\right>=i\hbar\frac{\partial\left|\Psi,t\right>}{\partial t} \tag{2.3}$$

   where $\hat{H}$ is the Hamiltonian (the energy operator) and $\left|\Psi,t\right>$ the normalized statevector at time t. Note that since the equation is of first order in time, knowledge of the statevector at the initial time t=0 uniquely determines it at any subsequent time.
6. Knowledge of the statevector yields all possible information one can have about an individual physical system (this assumption being referred to as the *completeness* of the theory), the predictions of the formalism being fundamentally probabilistic: for any conceivable observable $\hat{\Omega}$ the probability $P\left(\hat{\Omega}=\omega_k \mid \Psi,t\right)$ (or the probability density $P(\hat{\Omega}=\omega \mid \Psi,t)$) of getting the indicated outcome in a measurement of $\hat{\Omega}$ when the state is $\left|\Psi,t\right>$ is given by:

$$P(\hat{\Omega}=\omega_k\left|\Psi,t\right)=\left|c_k\right|^2=\left|<\omega_k\left|\Psi,t\right>\right|^2,\ (P(\hat{\Omega}=\omega\left|\Psi,t\right)=\left|c(\omega)\right|^2=\left|<\omega\left|\Psi,t\right>\right|^2) \tag{2.4}$$

7. When a measurement is performed and a specific result is obtained, wave packet reduction (WPR) takes place: the statevector of the system is transformed instantaneously into the normalized eigenvector associated to the obtained result (in the degenerate case in which there is a whole linear manifold associated to such a result - this typically occurring in the case of a result belonging to the continuous spectrum for which an infinitely precise measurement cannot be performed - the statevector is transformed into its normalized projection onto the considered linear manifold).

With reference to the axiomatic structure of the theory some remarks are appropriate:

I. If it happens that the statevector $\left|\Psi,t\right>$ coincides with one of the eigenstates $\left|\gamma_k\right>$ of an observable $\hat{\Gamma}$

$$\left|\Psi,t\right>=\left|\gamma_k\right>, \tag{2.5a}$$

then the probability of getting the result γ_k in a measurement of $\hat{\Gamma}$ equals 1 and one can legitimately state, following Einstein, Podolsky and Rosen (1935), that the system "objectively possesses" the considered property. Similarly, if $|\Psi, t>$ can be expressed as an integral of the eigenstates of the continuous spectrum confined to a certain interval Δ

$$\left(|\Psi, t>= \int_{\Delta} c(\gamma)|\gamma > d\gamma \right) \tag{2.5b}$$

the probability of getting the outcome belonging to the interval Δ equals one, so that one can state that the system objectively possesses the property $\hat{\Gamma} \in \Delta$. However, since in general the state is a superposition of eigenvectors belonging to different eigenvalus, e.g., $|\Psi, t>=\sum_k a_k |\gamma_k>$, there are nonepistemic probabilities $P(\hat{\Gamma}=\gamma_k |\Psi, t)=|a_k|^2$ of getting different outcomes. As a consequence one cannot claim that the system possesses an objective (i.e. independent of the measurement process) property related to $\hat{\Gamma}$.

II. It has to be stressed that in the considered case and if *completeness* is assumed, one cannot *even think* that the system possesses properties pertaining to $\hat{\Gamma}$ which are simply not known to the experimenter; in fact such an assumption would imply that there are ways of specifying more accurately the actual physical situation which would consent a precise prediction of the outcome. But this would amount to claim the incompleteness of the theory. Thus, in general, given a statevector and an observable, the situation corresponds to the system having *potentialities* referring to $\hat{\Gamma}$ which, however, require the act of measurement to become *actual*. This point must be always kept clearly in mind: the very logical structure of the theory implies that it makes probabilistic predictions about the outcomes of all conceivable measurement procedures, conditional on the measurement being performed.

III. If consideration is given to all conceivable observables of an individual physical system the very fact that they are associated to operators and not to functions as in the classical case, implies that they in general do not commute with each other. This in turn implies that (in general) an

eigenvector of an observable is a linear superpositions of eigenvectors corresponding to different eigenvalues for other observables. In brief, even if a property can be considered as possessed by the system, surely other properties are possessed only as potentialities. However, in the case of a system considered by itself (i.e. as isolated from the rest of the universe) there always exist at least an observable such that the statevector of the system is an eigenvector of the associated operator corresponding to a certain eigenvalue. Accordingly, the theory allows us to claim that any isolated system possesses objectively at least some properties. In this sense quantum mechanics can be considered as having taught us that one cannot attribute simultaneously too many properties to a given system. This is a simplified way of expressing Heisenberg's indeterminacy principle.

To conclude, we can concisely summarize the above analysis by stating that, according to quantum mechanics:

Some property is always possessed by an isolated system as a whole. However, quantum mechanics tells us that one cannot consider ***too many properties*** *as objectively possessed. There are* ***potentialities*** *that are actualized (by WPR) only if one actually performs an appropriate measurement.*

3. COMPOSITE SYSTEMS

The just outlined situation becomes much more complicated when one is interested in studying the constituents of a composite system. In such a case the already mentioned peculiar trait of quantum mechanics - entanglement (Verschrankung), in Schrödinger's words (Schrödinger, 1984): *the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical line of thought* - emerges and gives rise to further interpretational problems (p. 424). We stress that entanglement is a direct consequence of two precise and peculiar mathematical features of the formal language of the theory, i.e. of the fact that the Hilbert space of a composite system is the direct product of those of the constituents and of the linear character of the resulting Hilbert space itself.

For our purposes we can confine our considerations to the case of a system S made up only of two constituents: $S = S_1 + S_2$. In such a case two different types of states can be considered, the factorised and the entangled ones. The factorized states are simply the product of a state of one constituent times one of the other: $|\Psi>=|\Phi_k^{(1)}>\otimes|\Theta_j^{(2)}>$. For them, according to the previous discussion, there is an observable of system S_1 such that $|\Phi_k^{(1)}>$ is an eigenstate of the corresponding operator $\hat{\Omega}^{(1)}$ belonging, say, to the eigenvalue ω_k and an observable of system S_2 such

that $|\Theta_j^{(2)}>$ is an eigenstate of the corresponding operator $\hat{\Gamma}^{(2)}$ belonging, say, to the eigenvalue γ_j, so that we can state that the subsystem possesses the objective properties ω_k and γ_j. Obviously, other factorised states corresponding to different eigenvalues for both considered observables can be taken into account, e.g., a state $|\Xi>=|\Phi_r^{(1)}>\otimes|\Theta_s^{(2)}>$ with $\omega_r \neq \omega_k$ and $\gamma_s \neq \gamma_j$. But this is not the whole story. The linear nature of the state space of the theory tells us that if $|\Psi>$ is a possible state and $|\Xi>$ is also a possible state, then any linear (normalized) combination of them, in particular the entangled state:

$$|\Sigma>=\alpha|\Psi>+\beta|\Xi>, \quad |\alpha|^2+|\beta|^2=1, \tag{3.1}$$

is also a possible state for the system $S = S_1 + S_2$. Now, according to "the rules of the game" the theory tells us that, in such a state, the observable $\hat{\Omega}^{(1)}$ ($\hat{\Gamma}^{(2)}$) has probabilities $|\alpha|^2$ and $|\beta|^2$, respectively, of yielding the outcome ω_k (γ_j) or the outcome ω_r (γ_s) in a measurement. In other words the system has only potentialities concerning the considered observables $\hat{\Omega}^{(1)}$ and $\hat{\Gamma}^{(2)}$. Obviously, if a measurement of one (e.g. $\hat{\Omega}^{(1)}$) observable is performed a precise outcome (say ω_k) is obtained. Then WPR tells us that the state transforms instantaneously from $|\Sigma>$ to $|\Psi>=|\Phi_k^{(1)}>\otimes|\Theta_j^{(2)}>$, which, being an eigenstate of $\hat{\Gamma}^{(2)}$, implies that in a subsequent measurement of such observable, the outcome γ_j will be obtained with certainty.

It is important to point out that in the most general case it may even happen that the entangled state is not an eigenstate of any conceivable observable of the subsystems: the composite systems represents then an unbroken whole which, as such, has some property, but whose constituents have no properties at all. To clarify this point we will resort to an extremely simple case, i.e., we will limit our considerations to the spin space of a system of two spin-1/2 particles, and to the singlet state:

$$|\Psi(1,2)>=\frac{1}{\sqrt{2}}[|1,\uparrow>|2,\downarrow>-|1,\downarrow>|2,\uparrow>] \tag{3.2}$$

Actually, since the state is invariant for rotations, there is no need to specify the direction to which the arrows refer. Due to the fact that the arrows can point in any chosen direction, any spin measurement on one of the constituents has probability ½ of giving the outcome +1 and ½ of giving the outcome −1 (in units of $\hbar/2$). There follows, in particular, that there is no direction $\vec{n}$ such that:

$$P(\vec{\sigma}^{(1)} \bullet \vec{n} = +1 \,|\, \Psi) = 1 \tag{3.3}$$

This analysis focuses two extremely important facts about the composite system, i.e. that its constituents may have, in general, no actual but only potential properties, and, moreover, that a measurement performed on a constituent leading to the actualization of the measured quantity implies the instantaneous emergence of an actual property also for the other constituent, independently of the fact that the two constituents be far apart and non interacting (*quantum nonlocality*). Let us proceed.

4. INSTANTANEOUS OBJECTIFICATION AT-A-DISTANCE

Let us consider once more the singlet state and let us assume that S_1 and S_2 are far away and noninteracting. Let us decide to perform a spin measurement on one constituent, along an arbitrarily chosen direction. What does the theory tell us? To answer it is sufficient to write the state in the usual form with the arrow denoting the direction we have chosen:

$$|\Psi(1,2)>= \frac{1}{\sqrt{2}}[|1,\uparrow>|2,\downarrow> - |1,\downarrow>|2,\uparrow>] \tag{4.1}$$

Then:

- There is a probability ½ of getting the outcome +1 or −1 in the measurement
- According to the outcome, the statevector is reduced either to $|1,\uparrow\rangle|2,\downarrow\rangle$ or to $|1,\downarrow>|2,\uparrow>$,
- A subsequent measurement **along the same direction** of the spin of the other particle gives, with certainity, the opposite outcome.

  It has to be stressed that the measurement has **objectified** the measured quantity but, instantaneously, also the spin component along the same direction of the partner particle, in spite of the fact that it can be very far away and no more interacting with it. An element of physical reality for particle 2, which, if completeness is assumed, cannot even be thought to exist prior to the measurement, has emerged instantaneously at-a-distance.

5. THE SO-CALLED EPR PARADOX

We can now sketch the reasoning of the famous EPR-paper. It is based on the following assumptions:

- Completeness of the theory: ***there is nothing but the wave function***. Its knowledge represents the maximum possible specification of the state of an individual physical system. The quantum probabilities are then epistemic, i.e. they do not admit an ignorance interpretation.

- Reality criterion: *When one, without disturbing in any way the system, can predict the outcome of a prospective measurement with certainty then there is a **possessed property** or **element of physical reality** associated to the considered observable.*

- Einstein locality criterion: elements of physical reality cannot be influenced instantaneously at-a-distance .

The argument then goes as follows:

1. Before any measurement the theory attaches equal probabilities (1/2) to the two outcomes $\pm\hbar/2$ of any conceivable measurement of the spin component in any arbitrarily chosen direction $\vec{n}$: there is no objective element of physical reality, no property related to such observables.
2. One can choose to perform a measurement at one wing of the apparatus (let us say on system S_1): he gets one of the two outcomes, WPR induces the instantaneous transition from the singlet to the factorized state. Such a state attaches probability 1 to the outcome opposite to the one which has been obtained concerning the measurement of the spin component along the same direction $\vec{n}$ of the other particle. Immediately after the measurement a property of the far away particle has become objective.
3. Since immediately after the measurement on the system S_1 we can instantaneously predict with certainty the outcome of a measurement on system S_2, according to the locality requirement such a system must have possessed the property even before. However, there is no formal element of the theory that can be related to this fact: the theory is incomplete!

This conclusion of the EPR paper raises immediately an obvious question: is it possible to work out a deterministic (or even stochastic) completion of the theory in such a way to make epistemic the nonepistemic probabilities of the theory? In other words, the EPR paper led in a natural way to contemplate the possibility of a Hidden Variable Theory.

6. BELL'S INEQUALITY

At this point J.S. Bell enters the game. He, contrary to the adherent to the orthodox position about quantum mechanics, is seriously worried by the EPR argument. He knows that one can actually exhibit a deterministic completion of quantum mechanics, i.e., the de Broglie-Bohm pilot wave theory in its logically clean formulation due to Bohm (1952a, 1952b).

John Bell (1964) studies such a theory in all details and discovers that it has a characteristic feature: it is nonlocal in a quite precise sense. So he tries to find a local completion of quantum mechanics but he does not succeed. He gets the idea of proving that such a completion is impossible and derives his celebrated inequality.

The nice fact about his approach derives from its being based exclusively on an absolutely natural assumption of locality for space-like separated events. In order to appropriately specify its meaning one has to introduce some definitions. First of all, let us denote by λ the entities that specify the maximum possible knowledge that the theory allows about an individual physical system. Thus, within quantum mechanics with the completeness assumption λ must be identified with the statevector of the system, in a hidden variable theory with the hidden variables, in a theory like Bohmian mechanics with the statevector plus the positions of all particles of the system. To go on, let us specify the notation we will use. We denote as:

- $p_\lambda^{(1,2)}(\vec{a},\vec{b};\alpha,\beta)$ the probability of getting, for a given λ, the outcomes (=±1) α and β (=±1) in a joint measurement of $\vec{\sigma}^{(1)}\cdot\vec{a}$ and $\vec{\sigma}^{(2)}\cdot\vec{b}$,

- $p_\lambda^{(1)}(\vec{a},*;\alpha,*)$ and $p_\lambda^{(2)}(*,\vec{b};*,\beta)$ the probabilities of getting, for the same λ, the outcome α or β in a measurement of $\vec{\sigma}^{(1)}\cdot\vec{a}$ or $\vec{\sigma}^{(2)}\cdot\vec{b}$, respectively, when no measurement is performed on system S_2 or S_1.

We can now formulate in a mathematically precise way the locality assumption made by Bell which will be denoted as *B-Loc* by making reference to a situation in which the two measurements performed on the two subsystems of a pair of particles in the singlet state are space-like separated events:

$$B-Loc \Leftrightarrow p_\lambda^{(1,2)}(\vec{a},\vec{b};\alpha,\beta) = p_\lambda^{(1)}(\vec{a},*;\alpha,*)\bullet p_\lambda^{(2)}(*,\vec{b};*,\beta) \qquad (6.1)$$

Bell's locality requirement implies an inequality between appropriately defined quantities. We will not derive it (even though its proof is quite staightforward) but we will limit ourselves to state the result. For arbitrary directions $\vec{a}$ and $\vec{b}$, one defines the quantity:

$$E_\lambda(\vec{a},\vec{b}) = p_\lambda^{(1,2)}(\vec{a},\vec{b};+,+) - p_\lambda^{(1,2)}(\vec{a},\vec{b};+,-) - p_\lambda^{(1,2)}(\vec{a},\vec{b};-,+) + p_\lambda^{(1,2)}(\vec{a},\vec{b};-,-) \tag{6.2}$$

and shows that for any chosen directions $\vec{a},\vec{b},\vec{c},\vec{d}$ one has:

$$|E_\lambda(\vec{a},\vec{b}) - E_\lambda(\vec{a},\vec{d})| + |E_\lambda(\vec{c},\vec{b}) + E_\lambda(\vec{c},\vec{d})| \le 2 \tag{6.3}$$

which implies, denoting as $E(\vec{a},\vec{b})$ the average of $E_\lambda(\vec{a},\vec{b})$ over the hidden variables λ:

$$|E(\vec{a},\vec{b}) - E(\vec{a},\vec{d})| + |E(\vec{c},\vec{b}) + E(\vec{c},\vec{d})| \le 2 \tag{6.4}$$

Note that in accordance with its definitions and with the requirement that averaging on the possible hidden variables one gets the corresponding quantum predictions one has to identify $E(\vec{a},\vec{b})$ with the quantum mean value of the observable $\vec{\sigma}^{(1)} \cdot \vec{a}\, \vec{\sigma}^{(2)} \cdot \vec{b}$. For such a quantity standard Q.M. gives:

$$E(\vec{a},\vec{b}) = \cos[2\vec{a} \hat{\ } \vec{b}] \tag{6.5}$$

Then, if one chooses

$$\vec{a} = 0^0, \quad \vec{b} = 22.5^0, \quad \vec{c} = 45^0, \quad \vec{d} = 67.5^0$$

one gets:

$$|E(\vec{a},\vec{b}) - E(\vec{a},\vec{d})| + |E(\vec{c},\vec{b}) + E(\vec{c},\vec{d})| = 2 \times \sqrt{2} = 2.828 \tag{6.6}$$

i.e. quantum mechanics violates appreciably the predictions of any conceivable theory which satisfies the *B-Loc* condition.

The conclusion should be obvious: *completions of quantum mechanics (both deterministic and stochastic) are in principle possible (Bohmian*

Mechanics representing a paradigmatic example) but they are unavoidably nonlocal in the precise sense that they violate B-Loc.

7. EXPERIMENTAL METAPHYSICS

According to some authors, there were various loopholes in the experimental tests of Bell's inequality. The most relevant one was due to the fact that one chooses the directions along which to perform the spin measurements well in advance, so that there is, in principle, all the time in order that the choice of such directions allow a physical influence to propagate from one to the other wing of the apparatus. The quantum correlations might then be induced by a physical action from one wing of the apparatus to the other.

Alain Aspect (1982) accepted the challenge and, by taking advantage of important technological improvements, devised an experiment in which the choices of the orientations of the polarizers (he was working with photons) were made at a genuine space-like separation. He got agreement (within various standard deviations) with quantum predictions and a clear cut violation of Bell's inequality. Somebody still believes that the result is not conclusive by making appeal to the low efficiency of the detectors, but this way out is really untenable (in my opinion) by any serious scientist.

The now mentioned experiment has been appropriately denoted by Abner Shimony (1989) as an example of *Experimental metaphysics*, since it gives a clear cut experimental proof that one cannot think that one particle possesses objectively definite properties before the measurement on the other one is performed, but it acquires such properties instantaneously, in spite of the fact that the two measurements are space-like separated. To clarify the reasons for this position of Shimony one must remember that the Copenhagen orthodoxy was claiming that the reality request of Einstein (its pretension concerning the existence of objective properties of individual systems under appropriate circumstances) was due to his metaphysical prejudices concerning reality. In particular, since such a request could not be subjected to any experimental test, it was considered meaningless. It has been a merit of J.S. Bell to show how to test it and of Aspect to have performed the experiment.

8. THE CONCEPTUALLY REVOLUTIONARY IMPLICATIONS OF BELL'S INEQUALITY

Bell's fundamental result can be used to prove that, independently of any theory one adopts or of any conceivable interpretation of the formalism, if the experimental correlations between pairs of entangled particles in the singlet state predicted by quantum mechanics are true (i.e. if one ascertains that the experimental outcomes, agreeing with the predictions of quantum mechanics, violate Bell's inequality) then natural processes are basically nonlocal, something that nobody had suspected before the EPR paper and Bell's investigations.

Let us present a completely general proof of this statement. We first of all specify the notations we will use. We denote as

- $\{100\%Corr\}$: the fact that if one performs the same measurement at the two wings of the apparatus one always gets perfectly anticorrelated results.
- *Det*: the fact that all probabilities like $p_\lambda^{(1)}(\vec{a},*;\alpha,*)$, $p_\lambda^{(2)}(*,\vec{b};*,\beta)$ (and consequently, due to *B-Loc*, $p_\lambda^{(1,2)}(\vec{a},\vec{b};\alpha,\beta)$) can take only either the value 1 or the value 0, i.e. that the assignment of the hidden variables determines the precise outcome of any measurement.
- *Q.M.*: the assumotion athta all predictions of quantum mechanics, i.e. also the correlations concerning measurements in different directions at the two wings of the apparatus occur in accordance with the quantum predictions.
- *Bell's Ineq.*: the statement that Bell's inequality holds.

Before going on we need a lemma, i.e.: $\{100\%Corr\} \wedge \{B-Loc\} \supset Det$. Let us prove it.

Since *B-Loc* implies

i) $p_\lambda^{(1,2)}(\vec{a},\vec{a};+,+) = p_\lambda^{(1)}(\vec{a},*;+,*) \times p_\lambda^{(2)}(*,\vec{a};*,+) = 0$

ii) $p_\lambda^{(1,2)}(\vec{a},\vec{a};-,-) = p_\lambda^{(1)}(\vec{a},*;-,*) \times p_\lambda^{(2)}(*,\vec{a};*,-) = 0$

one has, either $p_\lambda^{(1)}(\vec{a},*;+,*) = 0$ or $p_\lambda^{(2)}(*,\vec{a};*,+) = 0$.

In the first case, since

iii) $p_\lambda^{(1)}(\vec{a},*;+,*) + p_\lambda^{(1)}(\vec{a},*;-,*) = 1$

iv) $p_{\lambda}^{(2)}(*,\vec{a};*,+) + p_{\lambda}^{(2)}(*,\vec{a};*,-) = 1$

one has $p_{\lambda}^{(1)}(\vec{a},*;-,*) = 1$. Then ii) implies $p_{\lambda}^{(2)}(*,\vec{a};*,-) = 0$ and iv) $p_{\lambda}^{(2)}(*,\vec{a};*,+) = 1$, and so on. Concluding: all elementary probabilities are then either 0 or 1 and the same holds for their products, i.e. *{Det}*.

We can now go on with our simple logical proof:

$$\{100\% Corr\} \wedge \{B - Loc\} \supset Det$$

$$Det \wedge \{B - Loc\} \supset Bell's\ Ineq.$$

$$Q.M. \supset \neg\{Bell's\ Ineq.\}$$

$$Q.M. \supset [\neg\{B - Loc\}] \vee [\neg Det]$$

$$[\neg Det] \supset [\neg\{100\% Corr\}] \vee [\neg\{B - Loc\}]$$

$$Q.M. \supset \{100\% Corr\}$$

$$Q.M. \supset \neg\{B - Loc\}$$

Thus, independently of any interpretation and with exclusive reference to the outcomes of our experiments we have to accept that natural processes are fundamentally nonlocal. Nobody, before Bell's theorem, had suspected that quantum processes, even though they were considered rather peculiar, might exhibit such unbelievable features.

9. CONCLUSION

It is important to stress that the identification of the extraordinary features analysed in the previous sections is a consequence of the specific behaviour of natural processes and of the fact that such processes require to resort to the Hilbert space language for their formalization. The linear nature of such a formalism and the fact that the description of composite systems demands the consideration of the direct product of the associated Hilbert spaces, as well as the fact that entangled states are natural elements of such a space, lead, by subtle formal and logical arguments, to the conclusion that most physical processes exhibit an unavoidable nonlocal character. I think this is a paradigmatic example of how the exploitation of the subtle mathematical properties of the formalism can lead to a deeper understanding of reality.

I hope to have been able to allow the reader to understand, with reference to the considered example, the fundamental role which mathematics plays for physics.

REFERENCES

Aspect, A., Grangier, P., and Roger, G., 1982, Experimental realization of the Eistein-Podolsky-Rosen-Bohm gedankenexperiment, *Phys. Rev. Letters*, **49**: 91-94.

Bell, J. S., 1964, On the Eistein-Podolsky-Rosen paradox, *Physics*, **1**: 195-200.

Bohm, D., 1952a, A suggested interpretation of the quantum theory in terms of 'hidden variables' I, *Phys. Rev.* **85**: 166-179.

Bohm, D., 1952b, A suggested interpretation of the quantum theory in terms of 'hidden variables' II, *Phys. Rev.* **85**: 180-193.

Einstein, A., Podolski, B., and Rosen, N., 1935, Can the quantum mechanical description of physical reality be considered complete?, *Phys. Rev.* **47**: 777-780.

Miller, A. I., 1987, *Imagery and Scientific Thought*, MIT Press, Cambridge, Massachusset.

Schrödinger, E., 1984, Discussion of probability relations between separated systems, in: *Gesammelte Abhandlungen – Collected Papers, Vol. 1*, Verlag der Osterreichischen Akademie der Wissenschaften, Wien, pp. 424-432.

Shimony, A., 1989, Search for a worldview which accommodate our knowledge of microphysics, in: *Philosophical Consequences of Quantum Theory*, J.T. Cushing and E. McMullin, eds., University of Notre Dame Press, Notre Dame, Indiana, pp. 25-37.

Wigner, E. P., 1959, The unreasonable effectiveness of mathematics in the natural sciences, in: 1995, *The Collected Works of Eugene Paul Wigner, Part B, Volume VI*, Springer Verlag, pp. 222-237.

THEORETICAL EXPLANATIONS IN MATHEMATICAL PHYSICS*

ANDRÉS RIVADULLA
Universidad Complutense, Madrid, Spain

Abstract: Many physicists wonder at the usefulness of mathematics in physics. According to Einstein mathematics is admirably appropriate to the objects of reality. Wigner asserts that mathematics plays an unreasonable important role in physics. James Jeans affirms that God is a mathematician, and that the first aim of physics is to discover the laws of nature, which are written in mathematical language. Dirac suggests that God may have used very advanced mathematics in constructing the universe. And Barrow adheres himself to Wigner's claim about the unreasonable effectiveness of mathematics for the workings of the physical world.

Wondering at the usefulness of mathematics in the physical description of reality is understandable indeed, if we assume that the laws, hypotheses, and theories of mathematical physics do describe, represent, or mirror Nature. But the fact that these physical constructs sometimes are empirically acceptable is no compelling logical reason for claiming that they do this job. The inference from empirical success to truth is logically illegitimate.

Theoretical models of physics use to be thought to represent reality. It is sometimes claimed that mathematical physics attempts to 'simulate' reality by means of models. But as the history of physics shows, it is perfectly possible to have different models of the same domain of phenomena, both empirically successful and based on entirely different assumptions. Thus theoretical models cannot be supposed to represent or simulate reality.

If instrumentalism about theories and theoretical models is adopted instead of realism in the philosophy of physics, the alleged unreasonable usefulness of

* This paper is part of a research, of reference BFF2002-01244, on theoretical models in physics, supported by the Spanish Ministry of Science and Technology.

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 161–178.

mathematics is less alarming. Mathematics then becomes an appropriate language, a useful instrument, in order to deal with Nature. But this has also consequences for the doctrine of theoretical explanations. Deprived of metaphysical connotations any physical construct, like facts, laws, hypotheses, etc., is considered to receive a theoretical explanation only when it has been deduced mathematically in the framework of another physical construct of higher level. Thus not only facts, but also empirical generalizations, abstract laws and even theories themselves admit of explanations in this sense. Now, explanation is, as well as prediction, the most important instance of realization of the hypothetic-deductive method. Since the methodology of physics is unthinkable without mathematics, mathematics becomes the possibility condition for theoretical explanations in physics.

Key words: mathematics; theoretical physics; realism; instrumentalism; empirical success; explanation; history of physics.

1. INTRODUCTION

Many physicists wonder at the usefulness of mathematics in physics. Allegedly mathematics is admirably appropriate for the description of natural phenomena, for the formulation of physical laws and theories. For instance Albert Einstein (1921) wonders at the 'enigma' of the certainty that mathematics gives the natural sciences:

> How can it be that mathematics, being after all a product of human thought which is independent of experience, is so admirably appropriate to the objects of reality? (p. 233)

Similarly, Paul Dirac (1963) claims

> It seems to be one of the fundamental features of nature that fundamental physical laws are described in terms of a mathematical theory of great beauty and power, needing quite a high standard of mathematics for one to understand it. You may wonder: Why is nature constructed along these lines? One can only answer that our present knowledge seems to show that nature is so constructed. We simply have to accept it. One could perhaps describe the situation by saying that God is a mathematician of a very high order, and He used very advanced mathematics in constructing the Universe.(p. 53)

Along with these lines Eugen Wigner (Wigner, 1967, p. 223) points to the problem, that, although mathematics often permits accurate descriptions of phenomena, we do not understand the reasons of its usefulness: "the enormous usefulness of mathematics in the natural sciences is something

bordering on the mysterious and …there is no rational explanation for it." Wigner (Wigner, 1967) concludes his article claiming that

> The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. We should be grateful for it and hope that it will remain valid in future research and that it will extend, for better or for worse, to our pleasure, even though perhaps to our bafflement, to wide branches of learning.(p. 237)

Among the philosophers of science the applicability of mathematics to reality has concerned for instance Gerhard Vollmer (Vollmer, 2001), who wonders: "How is it that mathematics, being silent about the world, can be used (so well) in the description of the world?". Also Vollmer refers to many scientists, some of them have been mentioned above, who make use of terms like 'riddle', 'secret', 'mystery' and 'miracle' in relation to this issue. Vollmer's response to this question is: Mathematics fits Nature, because

1) it is a structural science,
2) Nature is structured,
3) we are adapted to this structured world by evolution,
4) we are adapted to cognise some of these structures, and
5) we have language to devise non-mesocosmic structures.

This is a reasonable answer indeed. But it is incomplete, since it does not take into account a fact that Erhard Scheibe (Scheibe, 1992) has pointed to, as he also wondered at the amazing usefulness of mathematics, to wit: the *overdetermination* of physics by mathematics, i. e. the fact that we usually have in the physical theories more mathematics than we are able to interpret physically. According to Scheibe each physical law relates physical entities by means of mathematical operations for which there are no intended physical interpretations. Thus we can only enjoy the benefits of physically dealing with Nature at the costs of overdetermination, which seems to be unavoidable.

2. THE *NATURALNESS* OF THE USE OF MATHEMATICS IN PHYSICS

Wondering at the effectiveness of mathematics for the physical description of reality is not incompatible with the assumption of the naturalness of the use of mathematics in physics. For instance Albert Einstein (Einstein, 1933) maintains that "Our experience hitherto justifies us in believing that nature is the realization of the simplest conceivable mathematical ideas"(p. 274).

The naturalness of the use of mathematics in physics has been emphasized by many other physicists as well. For instance James Jeans, according to whom mathematical representations meet reality perfectly. Jeans (Jeans, 1954) claims that God is a mathematician, and that the first aim of physics is to discover the laws of nature, which are written in mathematical language. And the same applies to John D. Barrow (Barrow, 1992), who admits that the relationship between mathematics and physics is a symbiotic one. According to Barrow the reason of our success in revealing the working of the Universe is that we have discovered the language in which the book of Nature seems to be written, i.e. the language of mathematics. In spite of this Barrow (Barrow, 1992) maintains that

> There is one qualitative aspect of reality that sticks out from all others in both profundity and mystery. It is the consistent success of mathematics as a description of the workings of reality. (p. 173)

A success that is still considered by Barrow (Barrow, 1991, p. 175) the expression of an unreasonable effectiveness of mathematics in accounting for the workings of the physical world.

It is commonplace to point to Galilei's *Saggiatore,* 1623, in order to justify historically the effectiveness of the applicability of mathematics in physics: Book's Nature is written in mathematical language. But philosophers of science widely ignore that already nearly four centuries before Galilei, the founder of the Oxford's Franciscan school, Robert Grosseteste (1175-1253), had claimed[54] that

> *utilitas considerationis linearum, angulorum et figurarum est maxima, quoniam impossibile est sciri naturalem philosophiam sine illis.*

and that

> *omnes causae effectuum naturalium habent dari per lineas, angulos et figuras.*

Grosseteste's disciple Roger Bacon also maintained[55] that

> *impossibile est res huius mundi sciri, nisi sciatur mathematica.*

In the XIV century, the mathematization of Nature reached the climax among the *calculatores* at the Oxonian Merton College: Thomas Bradwardine (1290-1349), Roger Swineshead, and William Heytesbury (1313-1372), who for the first time posed the so-called *average velocity*

[54] *Cfr.* (Gilson, 1965, pp. 439-441).
[55] *Cfr.* (Gilson, 1965, pp. 444-449).

theorem: the space covered by an uniformly accelerated body is the same that he would cover with a constant velocity equivalent to the average value of its initial and final velocities. A theorem that repeatedly appeared in Nicolás de Oresme (1320-1382), Domingo de Soto (1494-1560) and Galileo Galilei (1564-1462).

But the history of the mathematization of Nature goes further back to the very beginnings of western science. Plato (427-347 BC) is not only the founder of mathematical astronomy. In the *Timeus* he also outlines a sort of mathematical protochemistry. Each atom of Empedocles' element theory: earth, air, water and fire, has an own geometric structure. The earth-atom is a hexahedron; the air-atom is an octahedron; the water-atom is an icosahedron, and the fire-atom is a tetrahedron.

With Plato starts geometric astronomy indeed. Plato's disciples Eudoxus and Calipus continued their master's astronomical way, making use of mathematical models intended *to save* celestial bodies' movements. It was the beginning of instrumentalism in physics and in the philosophy of science. The further development of astronomy, including the works by Ptolemy's, Copernicus, Brahe, Galilei and Kepler, until Newton, is branded by the question of whether mathematical models of the world do represent, mirror or describe reality as it is, or whether they merely save the observations. Thus the use of mathematics in astronomy is responsible for the polemic realism-instrumentalism in science, that has not abandoned the philosophical dispute ever since.

Any case it was obvious from the beginning of scientific theory that both celestial physics and mechanics would be unviable without mathematics. Mathematics was the *natural* way of dealing scientifically with Nature. This obviousness does not constitute any answer to the question of the effectiveness of mathematics in natural sciences, but it certainly makes this question less dramatic. Physics uses mathematics in order to build up models of reality and to formulate in the most precise way hypotheses about empirical phenomena.

3. INSTRUMENTALISM AND THEORETICAL EXPLANATIONS IN MATHEMATICAL PHYSICS

Wondering at the usefulness of mathematics in the physical description of reality is understandable indeed, if we assume that laws, hypotheses, and theories of mathematical physics do describe, represent, or mirror Nature. Theoretical models of physics use to be thought to represent reality. It is sometimes claimed that mathematical physics attempts to 'simulate' reality by means of models. But as Popper (Popper, 1983, 1994) has argued

correctly, models are vast and schematic oversimplifications, so that they simply cannot be true.

Mirroring reality is an illusory task, for, in order to picturing something, you need to have it in front of you, you have to know it. Otherwise you could not be sure that you are describing that portion of reality you intend to reflect. This is obviously not the common situation in physical sciences, where you only have some ways or appearances –usually not understandable at the beginning- by means of which an unknown phenomenon manifests itself. When we are able to construct theoretical models for phenomena, the fact that these physical constructs sometimes work is no compelling logical reason for claiming that they do simulate or represent Nature fairly. Therefore I disagree with Popper's view that we can determine by testing which models are nearer to the truth. But I reject as well that the relationships of models to reality is either isomorphism (Van Fraassen, 1980), or even the weaker views of similarity (Giere, 1988), or analogy (Boniolo, 2003). The inference from empirical success to truth, verisimilitude, similarity or analogy is logically illegitimate.

Finally as the history of modern physics shows, it is possible to have different models of the same domain of phenomena, both empirically successful and based on entirely different assumptions. This is the case for instance in nuclear physics[56], and in gravitational physics as well, where the successful Newtonian model based on forces and potentials was replaced by the geometrical model of general relativity theory, or by the hypothesis of gravitational interactions caused by exchange of gravitons, as quantum gravitation postulates. Thus theoretical models cannot be supposed to represent or simulate reality either.

If instrumentalism about theories and theoretical models is adopted instead of realism in the philosophy of physics, the alleged unreasonable usefulness of mathematics becomes less alarming: mathematics merely is an appropriate language, an useful instrument, in order to deal with Nature. And Scheibe's threatening doctrine of superdetermination of physics by mathematics loses a lot of its dramatic force.

This has serious consequences for the doctrine of theoretical explanations indeed. Deprived of metaphysical compromises I claim that any physical construct: facts, laws, hypotheses, etc., can only receive a theoretical explanation, when it can be deduced mathematically in the framework of another physical construct of higher level. Thus not only facts, but also empirical generalizations, abstract laws and even theories themselves admit

[56] As to the role played by theoretical models in nuclear physics see (Boniolo, 2002; Rivadulla, 2002*c*)

of explanations in this sense. Now, explanation is, as well as prediction, the most important instance of realization of the hypothetic-deductive method. Since the methodology of physics is unthinkable without mathematics, mathematics becomes the possibility condition for theoretical explanations in physics.

My viewpoint on theoretical explanations bases upon Einstein's assumptions on the hypothetic-deductive method of mathematical physics. Indeed, as to the theoretician's task, the search for explanations, Einstein claims (Einstein, 1914)

> The theorist's method involves his using as his foundation of general postulates or 'principles' from which he can deduce conclusions. His work thus falls into two parts. He must first discover his principles and then draw the conclusions that follow from them. […]
>
> Once this formulation is successfully accomplished, inference follows on inference, often revealing unforeseen relations that extend far beyond the province of the reality from which the principles were drawn. *But as long as no principles are found on which to base the deduction, the individual empirical fact is of no use to the theorist*; indeed he cannot even do anything with isolated general laws abstracted from experience. He will remain helpless in the face of separate results of empirical research, until principles that he can make the basis of deductive reasoning have revealed themselves to him. (p. 221, my italics)

Einstein insists on the idea of empirical laws waiting for explanation (Einstein, 1927):

> Newton's object was to answer the question: is there any simple rule by which one can calculate the movements of the heavenly bodies in our planetary system completely, when the state of motion of all these bodies at one moment is known? *Kepler's empirical laws* of planetary movement, deduced from Tycho Brahe's observations, confronted him, and *demanded explanation*. These laws gave, it is true, a complete answer to the question of *how* the planets move round the sun: the elliptical shape of the orbit, the sweeping of equal areas by the radii in equal times, the relation between the major axes and the periods of revolution. But these rules do not satisfy the demand for causal explanation. They are logically independent rules, revealing no inner connection with each other. (my italics)

This holds too for the separated laws that can be deduced from Planck's radiation law[57].

The ideal of a hypothetical-deductive physics is the following according to Einstein (Einstein, 1918)

> The supreme task of the physicist is to arrive at those universal elementary laws from which the cosmos can be built up by pure deduction. There is no logical path to these laws; only intuition, resting on sympathetic understanding of experience, can reach them. (p. 226)

Also in Einstein (Einstein, 1934):

> the grand aim of all science … is to cover the greatest possible number of empirical facts by logical deduction from the smallest possible number of hypotheses or axioms. (…) The theoretical scientist is compelled in an increasing degree to be guided by purely mathematical, formal considerations in his search for a theory, because the physical experience of the experimenter cannot lead him up to the regions of highest abstraction. The predominantly inductive methods appropriate to the youth of science are giving place to tentative deduction. Such a theoretical structure needs to be very thoroughly elaborated before it can lead to conclusions which can be compared with experience. Here, too, the observed fact is undoubtedly the supreme arbiter; but it cannot pronounce sentence until the wide chasm separating the axioms from their verifiable consequences has been bridged by much intense, hard thinking. The theorist has to set about this Herculean task fully aware that his efforts may only be destined to prepare the death blow to his theory. (p. 282)

At a first look, the object of the methodology of physics consists of accounting for facts or phenomena like: the retrogradation of planets, Mars' orbit, Balmer's spectral lines, the photoelectric effect, the *anomalous* Mercury's perihelion, natural radioactivity, etc. For instance Hanson claims that (Hanson, 1958)

> Phenomena are observed which are surprising and require explanation. […] The theoretician seeks concepts from which he can generate explanations of the phenomena. […[he aspires to fix the data in an intelligible conceptual pattern. (p. 123)

But from my viewpoint this is a rather restrictive way to look at explanation in physics. Further constructs are submitted to explanation too,

[57] *Cfr.* Rivadulla (2002*b*, pp. 152-154)

to wit: 1) empirical or phenomenological formulae like: Galilei's free fall laws, Kepler's laws of planetary motion, Stefan's law of black body's radiation, etc. But also 2) abstracts laws of theoretical physics, like: Boltzmann's distribution laws, Wilhelm Wien's radiation laws, etc., and 3) even theories themselves, are capable of receiving explanation, like the statistical justification of classical thermodynamics, or the mathematical deduction of Planck's radiation law in the framework of Bose-Einstein's quantum mechanical statistics.

Indeed, earlier or later physical constructs like empirical facts, phenomenological formulae, theoretical laws and even theories themselves become explained by more general laws and theories. When a physical construct can be deduced mathematically in the framework of a more general construct we affirm that it has received a *theoretical explanation*.

This use of explanation has nothing to do with metaphysical realism, i. e. it is not committed to any ontologically ready-made world expecting for intelligibility. Theoretical models and theories only are merely instruments enabling us to deal with Nature. No description or explanation of the independent world is possible. We live in a world that we cannot explain. Theoretical physics' main object is not to understand the world. No explanation in metaphysical sense is possible. Thus I disagree with Hanson, *op. cit., ibid.*, when he keeps on claiming:

When this is achieved he will know what properties fundamental entities do have; and he will have learned this by retroduction.

4. CASE STUDY I. THEORETICAL EXPLANATIONS OF THE HYDROGEN'S SPECTRAL LINES DISTRIBUTION

4.1 The *explanandum*: Johann Balmer's empirical formula

At the end of the XIXth century there was no explanation available for the spectra of the atomic elements. In spite of this fact, in 1885 the Swiss physicist Johann Jakob Balmer (1825-1898) found empirically, i. e. without reference to any theory about the atomic structure of matter, that the distribution of the Hydrogen spectral lines verifies that the number $\kappa = 1/\lambda$ is given by the formula:

$$\kappa = R_H \left(\frac{1}{4} - \frac{1}{n^2} \right),$$

where R_H=109677,576 cm^{-1} is known as the Rydberg constant, and $n \geq 3$. Thus it was urgent to account for both, the atomic spectra of elements and Balmer's formula.

4.2 Bohr's account of Balmer's empirical formula

In 1913 Niels Bohr (1885-1962) proposed his famous model of the Hydrogen atom. He assumed Rutherford's planetary model and completed it resorting to Planck-Einstein's quantum theory. Bohr's hydrogen atomic model was based on following postulates:

1. The electron is only allowed to have well defined energy values in stationary states.
2. In these stationary states the electron moves around the nucleus in circular orbits of radius r.

In order to maintain this postulate he introduced the bold hypothesis of the mechanical equilibrium among the Coulomb force, acting between atom nucleus and electron, and the centrifugal force due to the circular movement of the electron:

$$\frac{1}{4\pi\,\varepsilon_0}\frac{Z\,e^2}{r^2} = m\frac{v^2}{r}$$

3. When the electron jumps from an initial orbit to another orbit of lower energy, it does emit energy. The frequency of the emitted electromagnetic radiation by the electron is given by

$$\nu = \frac{E_i - E_f}{h}$$

where E_i and E_f denote respectively the energy of the initial and final orbits.

4. The *orbital angular momentum* of the electron is quantised, i. e. $L = mvr = n\hbar$.

From these postulates it follows that the total energy E of the electron – the sum of its kinetic energy E_k= ½mv^2 and its potential energy

$$V = -\frac{Ze^2}{4\pi\,\varepsilon_0\, r}$$ is

$$E = -\frac{m_e e^4 Z^2}{(4\pi\varepsilon_0)^2\, 2\hbar^2 n^2}$$

where $n = 1, 2, 3, \ldots$ is known as *Bohr's quantum number*.

We only need now to insert this value of the total energy in Bohr's expression

$$\nu = \frac{E_i - E_f}{h} = \frac{1}{2\pi}\frac{E_i - E_f}{\hbar},$$

and to take into account that the initial energy is less negative than the final one, in order to obtain

$$\nu = \frac{1}{(4\pi\varepsilon_0)^2}\frac{m_e e^4 Z^2}{4\pi\hbar^3}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right).$$

Now, since $\kappa = \frac{1}{\lambda} = \frac{\nu}{c}$, if we divide the expression above by c, we obtain

$$\kappa = \frac{1}{(4\pi\varepsilon_0)^2}\frac{m_e e^4 Z^2}{4\pi\hbar^3 c}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right).$$

Finally, as $R_\infty = \frac{1}{(4\pi\varepsilon_0)^2}\frac{m_e e^4}{4\pi\hbar^3 c} = 109737{,}3 cm^{-1}$,

widely agrees with Rydberg's constant, we conclude – the Hydrogen atomic number being Z=1- that

$$\kappa = R_\infty\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right).$$

In case n_f=2 we recover precisely *Balmer's formula*. Thus we are allowed to conclude that Bohr's atom model gives for the first time a *theoretical explanation* of Balmer's formula.

4.3 Erwin Schrödinger's account of Balmer's empirical formula

In spherical coordinates Schrödinger's time independent equation of hydrogenic atoms is[58]

$$-\frac{\hbar^2}{2\mu}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R\Theta\Phi}{\partial r}\right)+\frac{1}{r^2 sen^2\vartheta}\frac{\partial^2 R\Theta\Phi}{\partial\varphi^2}+\frac{1}{r^2 sen\vartheta}\frac{\partial}{\partial\vartheta}\left(sen\vartheta\frac{\partial R\Theta\Phi}{\partial\vartheta}\right)\right]+$$

$$+V(r)R\Theta\Phi = ER\Theta\Phi$$

where $R(r)\Theta(\vartheta)\Phi(\varphi)=\psi(r,\vartheta,\varphi)$ are its solutions, factorized in functions of only one coordinate. The application of the technique of variable separation allows solving Schrödinger's equation above in three steps:

(1) $$\frac{d^2\Phi}{d\varphi^2}=-m^2\Phi\,,$$

(2) $$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\frac{2\mu}{\hbar^2}\left[E-V(r)\right]R=\lambda\frac{R}{r^2}\,,$$

and

(3) $$\frac{m^2\Theta}{sen^2\vartheta}-\frac{1}{sen\vartheta}\frac{d}{d\vartheta}\left(sen\vartheta\frac{d\Theta}{d\vartheta}\right)=\lambda\Theta\,.$$

Equation (2) is known as *Schrödinger's radial equation*, and its solution allows us to obtain[59]

[58] *Cfr.* (Eisberg and Resnick, 1974).
[59] *Cfr.* (Eisberg and Resnick, 1974; Bransden and Joachain, 1983), etc.

$$E_n = -\frac{\mu Z^2 e^4}{(4\pi\varepsilon_0)^2 2\hbar^2 n^2},$$

where $\mu = (m\times M)/(m+M)$ is the reduced mass of the mass m of the electron and the mass M of the nucleus. Since the value just deduced of the electron energy is the same as the one obtained in the framework of Bohr's theory, we can legitimately claim that Balmer's empirical formula receives a theoretical explanation in the framework of the more sophisticated Schrödinger's wave mechanics. Moreover this allows us to conclude that Schrödinger's theory is empirically and theoretically more progressive than Bohr's atom theory.

5. CASE STUDY II. THEORETICAL EXPLANATIONS OF PLANCK'S RADIATION LAW

5.1 The *explanandum*: Max Planck's radiation law of the black body

Any body that absorbs all the radiation falling upon it is called a *black body*. Depending exclusively on its temperature a black body does radiate too. The radiation emitted by a black body in thermal equilibrium with the environment is called *black body radiation*, and its spectral distribution is the same for all black bodies, displacing itself to the shortest wave lengths with increasing temperature.

Wilhelm Wien (1864-1928) had obtained[60] that the energy density emitted by a black body satisfies the law

$$E(\nu,T) = \alpha\nu^3 e^{-\frac{\beta\nu}{T}}.$$

Max Planck (1858-1947) had also deduced that the energy density obeys the equation[61]:

[60] *Cfr.* (Jammer, 1989, pp.7-8)
[61] *Cfr.* (Jammer, 1989, Appendix A)

$$E(\nu,T)=\frac{8\pi\nu^2}{c^3}U\,,$$

where U denotes the average energy of an oscillator at temperature T. In October 1900 Planck was able to deduce that

$$U=\frac{b}{e^{b/T}-1}\,.$$

Thus Planck's *radiation law* took the form[62]:

$$E(\nu,T)=\frac{8\pi}{c^3}\frac{A\nu^3}{e^{B\nu/T}-1}\,,$$

where A and B are constants.

The problem with this law was that it was the result of a *glücklich erratene Interpolationsformel*. For a theoretician like Planck this law needed further theoretical justification.

5.2 Max Planck's account of Planck's radiation law

In order to recover the value of U on theoretical grounds Planck resorted to statistical thermodynamics. He thus assumed[63] 1) that S_N was the total entropy of a system of N oscillators of frequency ν and average energy U, and 2) that the total energy $U_N=NU$ of the system was equivalent to the energy $U_N= P\varepsilon$ of a whole number P of energy elements ε. Applying the second principle of statistical thermodynamics, $S_N=k\ln\Omega$, after a few calculations, Planck deduced that the value of U was:

$$U=\frac{\varepsilon}{e^{\varepsilon/kT}-1}\,.$$

Since *Wien's law* exiged that ε had to be proportional to ν, putting $\varepsilon=h\nu$ Planck obtained finally his famous *radiation law* for the black body:

[62] Published in Max Planck: "Über eine Verbesserung der Wienschen Spektralgleichung". *Verhandlungen der deutschen physikalischen Gesellsachft* 2, 1900, pp. 202-204. Reprinted in M. Planck, *Physikalische Abhandlungen und Vorträge*, Vol.1.

[63] *Cfr.* (Rivadulla, 2002*a*, pp. 53-54).

$$E(\nu,T)=\frac{8\pi\nu^2}{c^3}\frac{h\nu}{e^{h\nu/kT}-1},$$

which he presented[64] before the physical society on December 14th 1900. As it is very well known, this formula, which contained the hypothesis of the quantization of energy, started a new era in theoretical physics.

5.3 Bose-Einstein's quantum-mechanical account of Planck's radiation law

When a single photon is confined in a cubic box of side L, then his energy is given by

$$E=\frac{\hbar\pi c}{L}n\,.$$

It is very appropriate to take the radiation of a photon gas enclosed in a cavity as a model for the radiation of a black body. If the side L of the box is big compared with the radiation's average wave length, we can assume a continuum distribution of the energy of the photons. Moreover, we know from quantum mechanics that for every energy level of a particle there is a number g of different states. If we denote by $g(E)dE$ the number of states with energy in the interval $[E, E+dE]$, then, according to *Bose-Einstein distribution law*,

$$dN=\frac{g(E)dE}{e^{E/kT}-1}\,.$$

In order to compute the value of $g(E)dE$ we assume that the number $N(E)$ of states with energy E are the points of an sphere of radius n. In the volume of an octant of this sphere, the number of points will be[65]

$$N(E)=\frac{1}{8}\frac{4}{3}\pi n^3=\frac{1}{6}\pi n^3\,.$$

[64] He published it in Max Planck: "Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum". *Verhandlungen der deutschen physikalischen Gesellsachft* 2, 1900, pp. 237-245. Reprinted in M. Planck, *Physikalische Abhandlungen und Vorträge*, Vol.1.

[65] Basically I follow (Alonso and Finn, 1969).

From the value of energy above – knowing that L^3=V and that $\hbar = \dfrac{h}{2\pi}$ - we obtain $n^3 = 8V\dfrac{E^3}{c^3h^3}$.

Following

$$N(E) = \frac{8\pi V}{6}\frac{E^3}{c^3h^3}.$$

By differentiation,

$$\frac{dN(E)}{dE} = 4\pi V\frac{E^2}{c^3h^3}.$$

If we now call $dN(E)$=$g(E)dE$, and multiply by 2, due to the double polarization of photons, the number of states in the interval [E, E+dE] will be

$$g(E)dE = \frac{8\pi V}{c^3h^3}E^2 dE.$$

Obviously the number of states with frequency in [ν, ν+dν] will be

$$g(\nu)d\nu = \frac{8\pi V}{c^3}\nu^2 d\nu,$$

and, according to Bose-Einstein's distribution law, the number of photons with frequency [ν,ν+dν], is

$$dN = \frac{8\pi V}{c^3}\frac{\nu^2 d\nu}{e^{h\nu/kT} - 1};$$

multiplying the left side by E and the right side by hν and dividing the whole expression by V, we finally get

$$\frac{EdN}{V} = \frac{8\pi h\nu^3}{c^3}\frac{d\nu}{e^{h\nu/kT} - 1},$$

that is, Planck's radiation law of the black body

$$E(\nu) = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/kT} - 1}.$$

The mathematical deduction of Planck's radiation law amounts to its theoretical explanation in the framework of Bose-Einstein's statistical quantum mechanics.

REFERENCES

Alonso, M., and Finn, E. J., 1969, *Fundamental University Physics, Vol. III, Quantum and Statistical Physics,* Addison-Wesley, Reading, Massachusetts.

Barrow, J. D., 1992, *Perché il mondo è matematico*, Laterza, Roma.

Boniolo, G., 2003, Theories and models: Really old hat?, in: *Yearbook of the Artificial*,. Peter Lang Academic Publ. Co., Bern.

Boniolo, G., Petrovich, C., and Pisent, G., 2002, Notes on the philosophical status of nuclear physics, *Foundations of Science* 7: 425-452.

Bransden, B. H., and Joachain, C. J., 1983, *Physics of Atoms and Molecules*, Addison Wesley Longman Ltd, Essex, England, §§ 3.1 and 3.2.

Dirac, P. A. M., 1963, The Evolution of the Physicist's Picture of Nature, *Scientific American* 208 (5): 45-53.

Einstein, A., 1914, Principles of theoretical physics, Inaugural address before the Prussian Academy of Sciences, 1914. Reprinted in Einstein, A., 1973, *Ideas and Opinions*, Souvenir Press Ltd, London, pp. 221-223.

Einstein, A., 1918, Principles of research, Address delivered at a celebration of Max Planck's sixtieth birthday before the Physical Society in Berlin, 1918. Published in (1934) *Mein Weltbild*, Querido Verlag, Amsterdam. Reprinted in Einstein, A., (1973) *Ideas and Opinions*, Souvenir Press Ltd, London, pp. 224-227.

Einstein, A., 1921, Geometry and Experience, Lecture before the Prussian Academy of Sciences, January 27, 1921. Reprinted in Einstein, A., *Ideas and Opinions*, Souvenir Press Ltd., London, pp. 232-246.

Einstein, A., 1927, The mechanics of Newton and their influence on the development of theoretical physics, *Die Naturwissenschaften* 15. Reprinted in Einstein, A., 1973, *Ideas and Opinions*, Souvenir Press Ltd, London, pp. 221-223.

Einstein, A., 1933, On the method of theoretical physics, The Herbert Spencer Lecture, delivered at Oxford, June 10, 1933. Printed in Einstein, A., 1934, *Mein Weltbild*, Querido Verlag, Amsterdam. Second reprint in Einstein, A., 1973, *Ideas and Opinions*, Souvenir Press Ltd., London , pp. 270-276.

Einstein, A., 1934, The problem of space, ether, and the field in physics, in: *Mein Weltbild*, Querido Verlag, Amsterdam. Reprinted in Einstein, A., 1973, *Ideas and Opinions*, Souvenir Press Ltd, London , pp. 276-285.

Eisberg, R., and Resnick, R., 1974, *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles*, John Wiley and Sons, New York, §§ 7.3 and 7.4.

Giere, R., 1988, *Explaining Science. A Cognitive Approach*, University of Chicago Press, Chicago.

Gilson, É., 1952, *La philosophie au moyen âge*, Payot, Paris. Spanish version: 1965, *La filosofía en la Edad Media*, Gredos, Madrid.

Hanson, N. R., 1958, *Patterns of Discovery*, Cambridge University Press, Cambridge.

Jammer, M., 1989, *The Conceptual Development of Quantum Mechanics*, American Institute of Physics.

Jeans, J., 1954, *Physique et Philosophie*, Marcel Rivière et Cie, Paris, p. 97.

Popper, K. R., 1983, *Realism and the Aim of Science*, Hutchinson, London, pp. 58-59.

Popper, K. R., 1994, Models, instruments and truth, in: *The Myth of the Framework - In defence of science and rationality*. Routledge, London, p. 170 and ff.

Rivadulla, A., 2002a, La solución revolucionaria de Planck del problema de la radiación del cuerpo negro, in: *Física cuántica y realidad. Quantum Physics and Reality*, C. Mataix and A. Rivadulla, ed., Editorial Complutense, Madrid.

Rivadulla, A., 2002b, The search for explanations in the methodology of mathematical physics, in: *Science, Philosophy and Music*, E. Neuenschwander and L. Bouquiaux, ed., Proceedings of the XXth International Congress of History of Science, Vol. XX, Brepols Publishers, Turnhout, Belgium.

Rivadulla, A., 2002c, Representación por modelos y éxito empírico en ciencia", in *Semana del Pensamiento Filosófico. Popper: Los grandes debates del Siglo XX*. Universidad Industrial de Santander, Bucaramanga, Colombia. Año 3, No.3, Octubre 2002: 105-117.

Scheibe, E., 1992, The role of mathematics in physical science, in: *The Space of Mathematic*, J. Echeverria *et al.*, ed., *Philosophical, Epistemological and Historical Explorations*. Walter de Gruyter, Berlin.

Van Fraassen, B., 1980, *The Scientific Image*, Oxford University Press, Oxford.

Vollmer, G., 2001, Why does mathematics fit nature? The problem of application, in: *Proceedings of the III-IV International Ontology Congress*, Gómez Pin, V., ed.,. *Ontology Studies/Cuadernos de Ontología* Nº 1-2, 2001: 301-309

Wigner, E. P., 1967, The unreasonable effectiveness of mathematics in the natural sciences, in: E. Wigner, *Symmetries and Reflections*, Indiana University Press, Bloomington, Indiana, pp. 22-237. Reprinted in: E. Wigner, *Philosophical Reflections and Syntheses*, Springer, Berlin 1995, pp. 534-549.

MATHEMATICS, PHYSICS AND MUSIC
A Case Study

ZVONIMIR ŠIKIĆ
University of Zagreb, Zagreb, Croatia

Abstract: I discuss the Pythagorean law of small numbers and its use in interpretations of our sensory discriminations of consonance vs. dissonance. It seems that the fact of non-western musical traditions contradicts the law and forces us to interpret the discriminations as acquired and subjective. I would like to show that this is a wrong interpretation, because it is based on the irrelevant empirical evidence. It does not take into account the correct mathematical and physical explanation of the law, provided by Helmholtz's theory in 1877 and corroborated by Plomp-Levelt experiment in 1965.

Key words: consonance; dissonance; Pythagorean law of small numbers; Helmholtz's dissonance curve; Plomp-Levelet experiment.

1. THE PROBLEM

The Pythagoreans came to believe that principles of mathematics are the principles of everything. The starting point of this rather general belief was their discovery of "*the law of small numbers*" i.e. their discovery that the pitch of a string is simply related to its length. When the length is shortened in ratio 1:2 the pitch jumps up an octave, when shortened in ratio 2:3 it jumps up a fifth, in ratio 3:4 it jumps up a fourth, in ratio 4:5 a major third etc.

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 179–196.

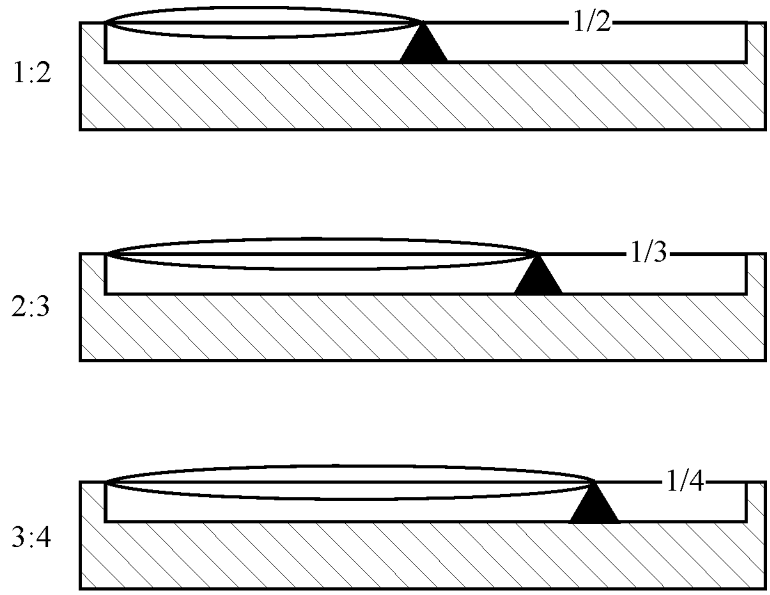


Figure 1. Ratio 1:2 produces an octave, 2:3 produces a fifth and 3: 4 a fourth.

To shorten the length is to enlarge the frequency and we could say that the Pythagoreans discovered that the frequency ratio between the octave and the fundamental is 2:1, between the fifth and the fundamental 3:2, between the fourth and the fundamental 4:3 etc.

Pythagoreans proceeded to describe the whole universe in terms of simple harmonic relationships; from the harmonious or inharmonious resonances in human bodies bellow the moon, to the harmony of the spheres above. To use the nomenclature of a later era, *musica instrumentalis*, the ordinary music made by plucking the lyre, was extended from *musica humana* to *musica mundana*.

What interests us here is the following question. Do our discriminations of consonant and dissonant intervals have some basic origin in facts “out there” in the real world?

According to the law of small numbers it seems that there is something *unique* “out there” which we discriminate as consonance “in here”. This *unique* source of our discriminations is the harmonic sequence of frequencies $1f : 2f : 3f : 4f : \dots$. This sequence of integer frequencies is different from all the other non-integer sequences and we perceive this *objective* difference as consonance. In this sense our discriminations are *objective* and not subjective.

This objective explanation of the consonance discriminations is the common opinion in western arts and sciences. We illustrate it with a quotation from 17th century scientist:

> The laws of music are unchangeably fixed by nature, hence they should hold not only for the entire earth, but fot the inhabitants of other planets as well. (C. Huygens as quoted in (Perlman, 1994)).

and a 20th century artist:

> A music – whether folk, pop, …, tonal, atonal, …, past, future, … - all of it has a common origin in the universal phenomenon of the harmonic series. (Bernstein, 1976).

But there is a huge problem with the common opinion. It is the existence of non-western musical traditions whose consonant intervals have nothing to do with the harmonic series. For example, the gamelan percussive orchestra, which is the indigenous musical tradition of Java and Bali, use 5 tone *slendro* and 7 tone *pelog* scales. Neither scale lies even remotely close to the western harmonic scales. Their consonances are based on non-integer sequences of frequencies.

Hence, there is *nothing unique* “out there” which humans discriminate as consonances “in here”. It seems that our discriminations are *subjective* and not objective.

We have two opposing conclusions. According to the law of small numbers Pythagorean just intonation, which is based on the integer sequence of frequencies, is a human universal. If we take into account the existence of the non-western musical traditions, whose scales are based on many different non-integer sequences of frequencies, it is not a human universal[66].

2. ANOTHER DIMENSION

There is also another dimension of the problem. Do our discriminations depend on innate systems, as we have tacitly presupposed until now, or do they depend on our experiences? In other words, are these discriminations *innate* or *acquired*[67]? According to the law of small numbers it seems that they are innate and objective. According to the fact of the existence of

[66] Note that it does not necessarily mean that intonation is not human universal. (English language is not human universal although language could be)

[67] By innate I mean acquired by evolution at least in some respects. By acquired I mean acquired exclusively by culture.

different musical traditions they could be acquired and subjective or perhaps innate and subjective[68]. Prevailing opinion is that they are acquired and that means changeable. Around this opinion evolved a lot of music-policy nonsense:

Musical racist imperialism:
The music of other cultures should evolve towards western higher forms which are based on immutable laws of nature.

Musical cultural imperialism:
The music of other cultures should evolve towards OUR higher forms which are produced by OUR superior culture.

Musical cosmopolitanism:
All musical traditions are equally worthy and should influence each other

Musical nationalism:
It is OUR music and we do not want any influences.

As far as it is based on the notion of the consonant intervals it is all wrong, because it is based on the irrelevant empirical evidence. In particular it does not take into account what happened to the law of small numbers in the last few thousand years and to the understanding of the other musical traditions in the last century. Let me explain.[69]

3. GALILEO'S THEORY

Notice that Pythagoreans offered no explanation of the law of small numbers. To offer one you should have some ideas about sound.

If you focus on perceptual aspect, sound is the sensation stimulated in the organs of hearing by vibrations in the air with frequencies in the range of 20 to 20 000 Hz. The vibrations are vibrations of a pressure wave, also known as a sound wave. It is explained in the following figure.

[68] Let me show you with some examples that our auditory discriminations can be of any of the four types. Our discrimination between loud and soft sound is innate and objective; between a string and a wind instrument it is acquired and objective; between an ugly and a beautiful piece of music it is acquired and subjective; between the mother tongue and a foreign language it is innate and subjective (cf.[1)]).

[69] My explanation follows (Sethares, 1997).

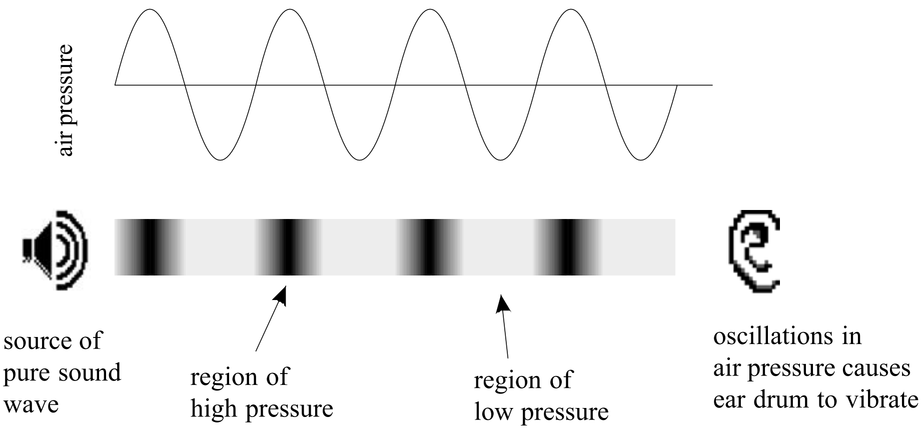


Figure 2. Vibrations of a pressure wave are perceived as sound.

The peaks represent times when air molecules are clustered, causing higher pressure. The valleys represent times when the air density, and hence the pressure, is lower. The wave pushes against the eardrum in times of higher pressure, and pulls during times of low pressure, causing the drum to vibrate. These vibrations are perceived as sound.

In accordance with this general idea Galileo offered one of the first explanations of the law of small numbers (Galilei, 1974)

> … agreeable consonances are pairs of tones which strike the ear with a certain regularity; this regularity consists in the fact that the pulses delivered by the two tones, in the same interval of time, shall be commensurable in number, so as not to keep the eardrum in perpetual torment, bending in two different directions in order to yield to the ever discordant impulses.

Galileo's pulses are the periods of the corresponding sound waves. If we represent them as below, then the number of points per unit interval represents the corresponding frequency.

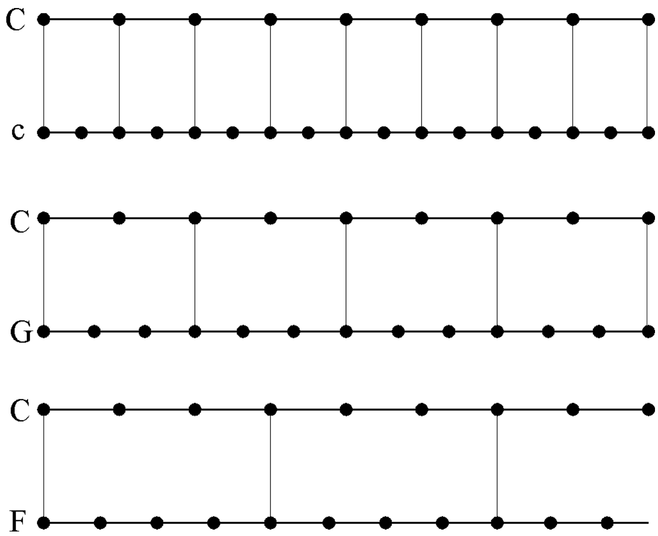


Figure 3. A representation of Galileo's pulses.

Looking at this representation we could say that one half of the octave is contained in the fundamental, and that explains the intimacy of the octave and its fundamental. Similarly, one third of the fifth is contained in the fundamental, and that explains a bit less intimacy of the fifth and its fundamental. In the same way one fourth of the fourth is contained in the fundamental, one fifth of the major third, one sixth of the minor third etc. That explains their diminishing consonances.

The same pattern could be represented arithmetically.

Table 1. One half of the octave is contained in the fundamental.

C		2		4		6		8		10		12	…
c	1	2	3	4	5	6	7	8	9	10	11	12	…

Table 2. One third of the fifth is contained in the fundamental.

C		3		6		9		12		15		18	
G	2		4	6	8		10	12	14		16	18	…

Table 3. One fourth of the fourth is contained in the fundamental.

C		4		8		12		16		20		24	…
F	3		6		9	12	15		18		21	24	…

Note how initial ratios determine the intimacy of the tones in given intervals i.e. their consonances.

Galileo's theory is very nice and frequently cited, even today, but there is one big problem with it. It is not true.

4. THE TRUE THEORY FOR SIMPLE SOUNDS

In an important experiment in 1965, Plomp and Levelt investigated how untrained listeners judge the dissonance[70] of a variety of intervals *when sounded by pairs of pure sine waves*. The result of the experiment is represented by the dissonance curve.

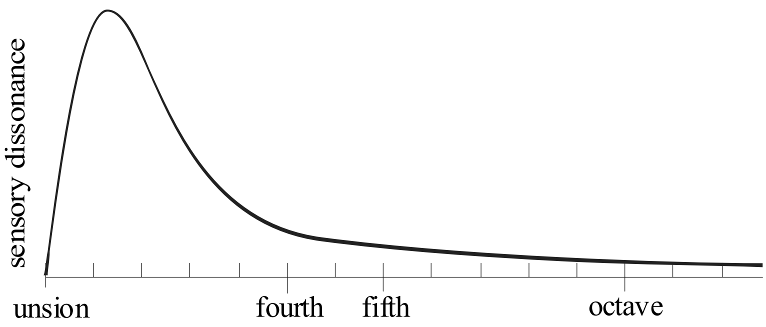


Figure 4. The dissonance curve of Plomp and Levelet.

(1) The dissonance is minimum, zero, when both sine waves are of the same frequency.

(2) It increases rapidly to its maximum somewhere around the second, in the middle range.

(3) Then it decreases steadily back toward zero.

Notice that major 7th and minor 9th are almost indistinguishable from the octave in terms of sensory dissonance for pure sine waves. This is in complete disagreement with Galileo's theory.

[70] The dissonance was defined as unpleasantness.

Helmholtz explained what is happening here, almost a century before Plomp and Levelet made their experiment.[71] He based his explanation on the phenomenon of beating which we explain very briefly.

The phenomenon of beating is caused by alternation of constructive and destructive interference. When two sine waves of exactly the same frequency are played together they sound just like a single wave, but the combination may be louder or softer then the original waves. When the waves have the same phase, the same starting point, their peaks and valleys line up exactly and the magnitude of the sum is greater then either are alone. This is constructive interference. When the waves which are out of phase are added together the peaks of one could line up with the valleys of the other and their sum is smaller then either alone. This is destructive interference.

What if the two sine waves differ slightly in frequency? The easiest way to picture this is to imagine that the two waves are at the same frequency, but that their relative phase slowly changes. When the phases are aligned they add constructively, while when out of phase they add destructively. The result is beating.

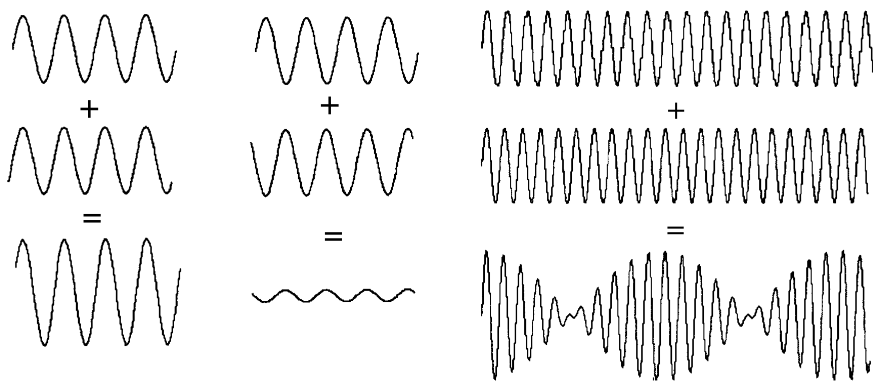


Figure 5. Constructive interference. Destructive interference. Beating.

Now, it is easy to understand the already announced Helmholtz's explanation:

1. When the sine waves are very close in frequency they are heard as a simple pleasant tone with slow vibrations in loudness. The physical origin of this pleasant vibrato is the phenomenon of beating.

[71] Helmholtz could make a relevant experiment only with complex sounds produced by then available instruments. The pure sounds, so easy available on computers these days, were not so easily available in Helmholtz's days.

2. Somewhat further apart in frequency the beating becomes rapid and this is heard as dissonance.
3. Then the tones separate and are perceived individually as a consonant pair.

   It is illustrated in the following figure.

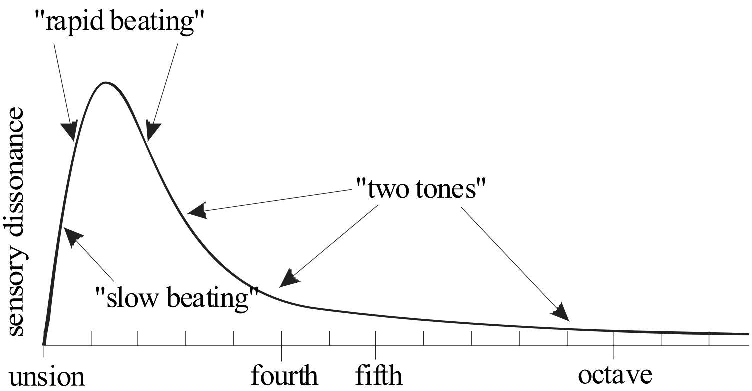


Figure 6. Helmholtz's explanation of sensory dissonance.

5. COMPLEX SOUNDS

We are really interested in complex sound waves produced by our musical instruments and not in the pure sine waves. The pure sine waves are important only because the complex waves are made of them.

Just as a complex light wave is made of the rainbow spectrum of pure color waves, a complex sound wave is made of pure sine waves in various bass, midrange and treble frequencies. First could be analyzed by a prism, second by the Fourier analysis.

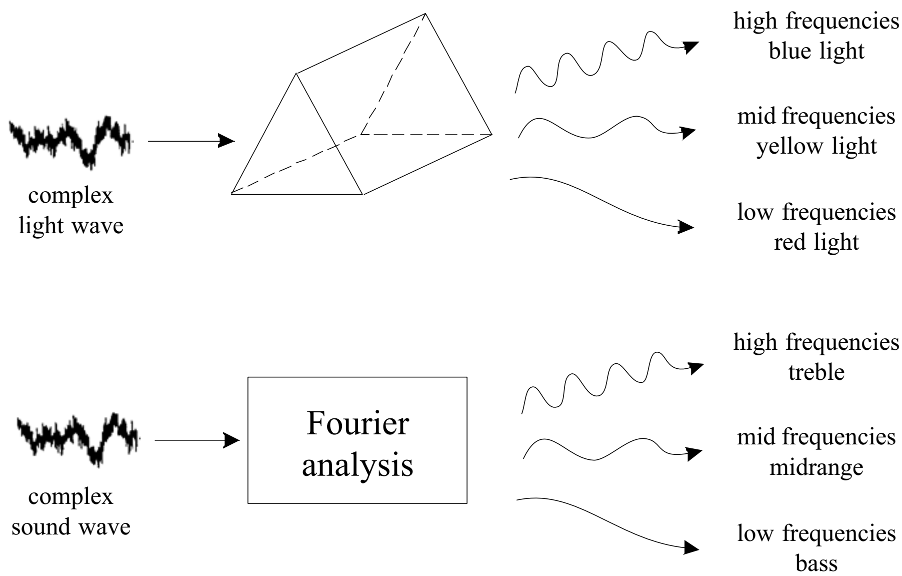


Figure 7. The analysis of the complex light and sound waves.

The Fourier analysis reduces a complex sound wave to its spectrum of frequencies. For example, the complex sound waves (d) and (e), which are (a) + (b) and (a) + (c) respectively, are both reduced to spectrum (f), which rediscovers the frequencies of the original sine waves.[72]

[72] The figure is from (Sethares, 1997), p. 15.

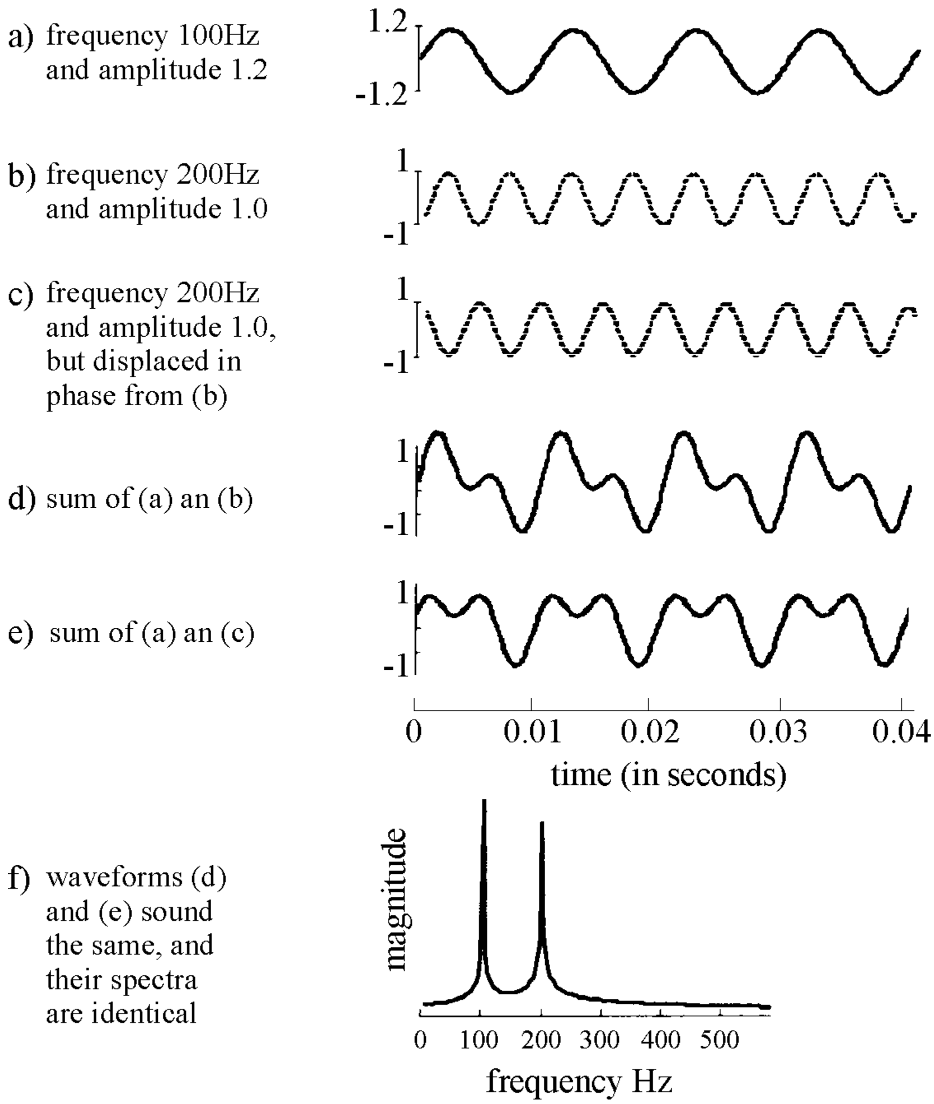


Figure 8. What we hear is the spectrum of a sound wave.

Our auditory system is a biological spectrum analyzer doing the same. It transforms a sound wave into a frequency spectrum which has an auditory meaning. (G. Ohm was first to propose this idea in 1843.) This is explained in the following illustration.[73]

[73] Ibid. p.16.

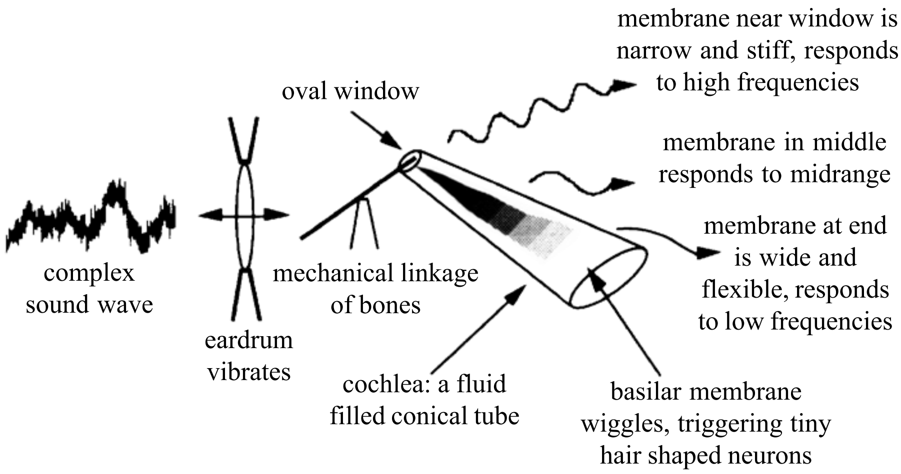


Figure 9. Our auditory mechanism.

The vibrations are transferred to the cochlea[74] which is filled with fluid. The motion of the fluid rocks the membrane spread along the cochlea. The region nearest the oval window responds to high frequencies, while the far end responds to low frequencies. Tiny neurons sit on the membrane sending messages towards the brain when jostled.

Thus the ear takes in a sound wave, like (d) or (e) above, and sends to the brain a representation of its spectrum, like (f) above. This representation has an auditory meaning.

6. HARMONIC AND NON-HARMONIC SOUNDS

As we said above, what really interests us is how to explain the dissonance of variety of intervals, when sounded by pairs of *complex* sound waves. These are sounds produced by our musical instruments.

First of all there is a big difference between complex sounds that are *harmonic* and those that are *not harmonic*. We introduce these two kinds of sound with two examples.

A typical example of the harmonic sound is the sound of a guitar pluck. Here is its spectrum.

[74] The cochlea is straightened out in the illustration. In reality it is curled up like a snail shell.

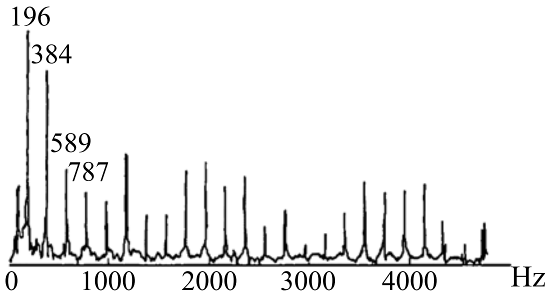


Figure 10. The spectrum of a guitar pluck.

Notice that the spectrum consists of the fundamental at $f = 196$ Hz and of the near *integer* partials at $2f \approx 384$ Hz, $3f \approx 589$ Hz, $4f \approx 787$ Hz etc. Such a spectrum in which all the frequencies of vibration are integer multiples of some fundamental f is called *harmonic* and the corresponding sound is called *harmonic sound*. Since every partial repeats exactly within the period of the fundamental, harmonic sound waves are periodic.

A typical example of the non-harmonic sound is the sound of the strike of a metal bar. Here is its spectrum.

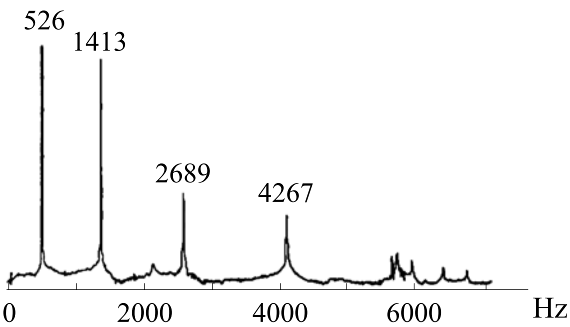


Figure 11. The spectrum of the strike of a metal bar.

Notice that the spectrum consists of the fundamental at $f = 526$ Hz, and of the *non-integer* partials at $2.68 f = 1413$ Hz, $5.11f = 2689$ Hz and $8.11 f = 4267$ Hz. Such a spectrum in which the frequencies of vibration are not the integer multiples of some fundamental f is called *non-harmonic* and the corresponding sound is called *non-harmonic sound*. Since at least some partials do not repeat exactly within the period of the fundamental, harmonic sound waves are not periodic.

The guitar string and the metal bar are only two of many possible sound making devices. The harmonic vibrations of the string instruments are also characteristic of many other musical instruments. For example, when air

oscillates in a wind instrument, its motion is constrained in the same way that the string is constrained by its fixed ends. At the closed end of the wind instrument the flow air must be zero, while at an open end the pressure must drop to zero. Thus all wind and string instruments have spectra which are harmonic. In contrast, most percussion instruments such as drums, marimbas, gongs etc. have non harmonic spectra.

The spectrum of the string is harmonic because the string is fixed at both ends, and can only sustain oscillations that fit exactly into the length of the string. It is possible to prove mathematically that for an ideal string, if the fundamental occurs at frequency f, the second partial must be at $2f$, the third at $3f$ etc.

The spectrum of the metal bar is non-harmonic because the bar is free at both ends. Hence, the movement of the struck bar is characterized by “bending modes” that specify how the bar will vibrate once it is set into motion.

It is possible to prove mathematically that for an ideal metal bar, if the fundamental occurs at frequency f, the second partial must be at $2.76\,f$, the third at $5.41\,f$, the fourth at $8.94\,f$ etc.

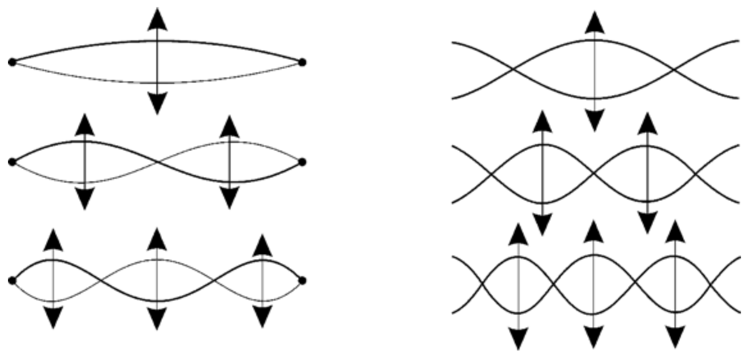


Figure 12. The "bending modes" of the string. The "bending modes" of the metal bar.

7. THE TRUE THEORY FOR COMPLEX SOUNDS

Let us return to our main question. How to explain the dissonance of variety of intervals, when sounded by pairs of *complex* sound waves?

The Plomp-Levelt experiment gathered data only on perceptions of pure sine waves. A century before that, to explain the sensory dissonance of complex sounds, Helmholtz proposed the following procedure: *add up all of the dissonances, between all pairs of pure sine partials*.

Notice that even if there is no beating interference of the fundamentals, there can be some beating interference of the other partials. Here is one example:

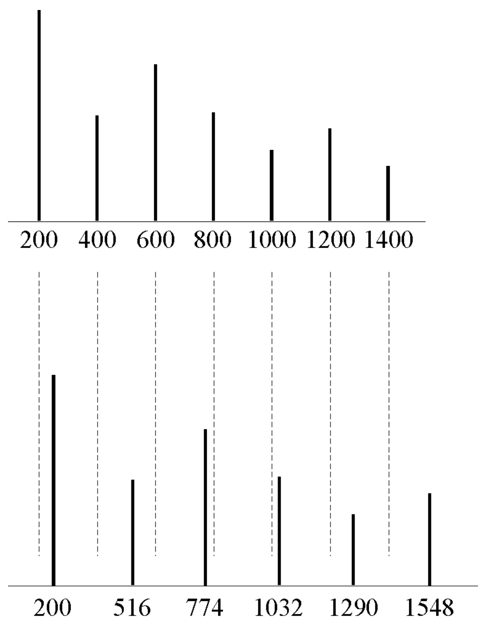


Figure 13. Some partials are interfering although the fundamentals are not.

A harmonic sound at fundamental frequency $f = 200$ Hz is transposed to $g = 258$ Hz. When this interval is played simultaneously some of the partials interfere by beating rapidly, causing sensory dissonance.

If we add up dissonances between all pairs of partials for all intervals we will get the dissonance curve for a given spectrum.

The dissonance curve for a harmonic spectrum with six partials at f, $2f$, $3f$, $4f$, $5f$ and $6f$ is shown in the following figure.[75]

[75] Ibid. p. 92. This is the figure that Helmholtz got as the result of his calculation.

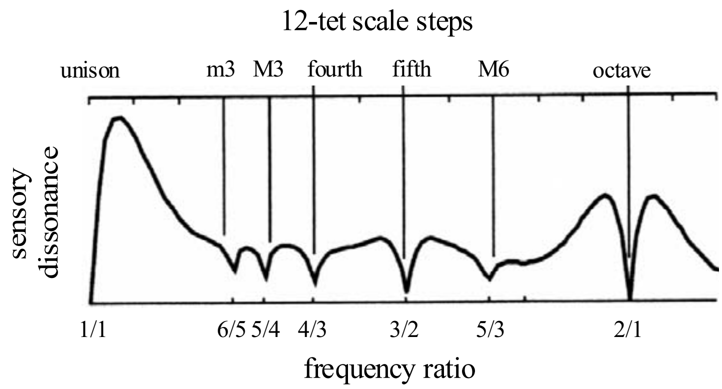


Figure 14. The dissonance curve for complex harmonic sound.

Notice, that minima of the dissonance curve coincide with many Pythagorean intervals, which are characterized by the law of small numbers. It is easy to prove that dissonance curves of harmonic spectra always have this property, and this is the *final explanation of the law of small numbers* for harmonic sounds.

The dissonance curve for a non harmonic spectrum of a metal bar with six partial at f, $2.76f$, $5.41f$, $8.94f$, $13.35f$ and $18.65f$ is shown in the following figure[76].

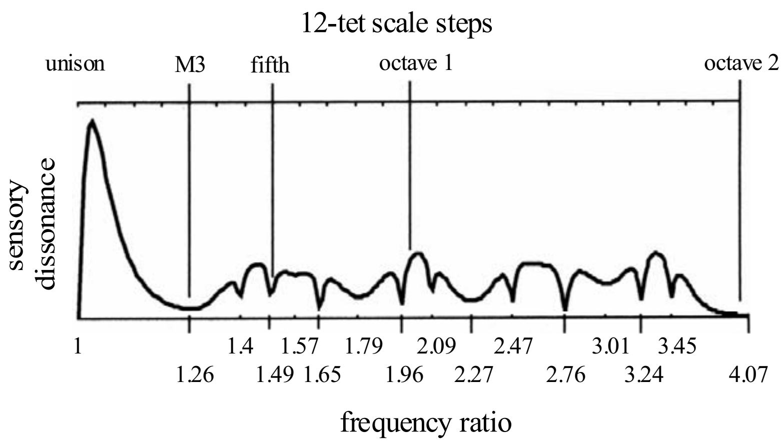


Figure 15. The dissonance curve for complex non harmonic sound.

[76] Ibid. p. 107. Helmoltz did not make the calculations for non-harmonic spectra because he was focused exclusively on harmonic instruments.

Notice, that minima of the dissonance curve do not coincide with any of the Pythagorean intervals, which are characterized by the law of small numbers. It is easy to prove that dissonance curves of non-harmonic spectra always have this property, and this is the *final refutation of the law of small numbers* for non harmonic sounds.

8. CONCLUSION

We may conclude that the law of small numbers is just an epiphenomenon which is empirically irrelevant for our explanations as much as are the non-integer "laws" of slendro, pelog and other non-harmonic scales. The real source of our consonance discriminations is the phenomenon of beating, as hypothesized by Helmholtz and directly corroborated by Plomp-Levelt experiment[77]. The beating is really something *unique* "out there" which we discriminate as dissonance "in here" and in this sense our discriminations are *objective* and not subjective. Furthermore, it is common to all musical traditions harmonic or not to discriminate between consonant and dissonant intervals in this way. It seems then that this is common to all humans, which means that our consonance discriminations are innate and not acquired. Hence, our sensory discriminations of consonant and dissonant musical intervals are *objective* and *innate* and this is corroborated by the western harmonic tradition as well as by the non-western non-harmonic traditions.

To be more specific we may say that sensory dissonance and consonance are functions of the interval *and the spectrum of the sound.* A scale and a spectrum are related if the dissonance curve for the spectrum has minima at the scale steps. Harmonic spectra of western musical instruments are related to western scales with many Pythagorean intervals. Non-harmonic spectra of different musical traditions are related to their scales.

And this is not the whole story. Nowadays musicians compose for very unusual sounds. In accordance with the previous explanations, their procedure is as follows:

(1) Choose a sound.

(2) Find the spectrum of the sound.

(3) Simplify the spectrum.

[77] It was indirectly corroborated by Helmholtz when he calculated that the minima of the dissonance curve for harmonic sounds (he was exclusively dealing with) correspond to Pythagorean intervals of the western harmonic tradition.

(4) Calculate the dissonance curve.

(5) Choose a set of intervals from the minima i.e. choose related scale.

(6) Create (synthesize) an instrument with the simplified spectrum that can play the sound at the chosen scale steps.

(7) Compose and play music.

We made the full circle. From music, to the first empirical laws, to their mathematical refinements and physical corroborations, and finely back to the music. To appreciate that you should listen to some music composed according to the procedure (1)–(7), which is the by product of this full circle history. The best starting point I can suggest is (Sethares, 1997).

REFERENCES

Bernstein, L., 1976, *The Unanswered Question*, Harvard University Press.

Galilei, G., 1974, *Two New Sciences*, transl. by S. Drake, University of Wisconsin Press.

Perlman, M., 1994, American gamelan in the garden of eden: intonation in a cross cultural encounter, *Musical Quaterly* **78**: 510-555.

Sethares, W. A., 1997, *Tuning, Timbre, Spectrum, Scale*, Springer.

THEORETICAL MATHEMATICS

On the Philosophical Significance of the Jaffe-Quinn Debate

MICHAEL STÖLTZNER
University of Bielefeld, Bielefeld, Germany

Abstract: Answering to the double-faced influence of string theory on mathematical practice and rigour, the mathematical physicists Arthur Jaffe and Frank Quinn have contemplated the idea that there exists a 'theoretical' mathematics (alongside 'theoretical' physics) whose basic structures and results still require independent corroboration by mathematical proof. In this paper, I shall take the Jaffe-Quinn debate mainly as a problem of mathematical ontology and analyse it against the backdrop of two philosophical views that are appreciative towards informal mathematical development and conjectural results: Lakatos's methodology of proofs and refutations and John von Neumann's opportunistic reading of Hilbert's axiomatic method. The comparison of both approaches shows that mitigating Lakatos's falsificationism makes his insights about mathematical quasi-ontology more relevant to 20^{th} century mathematics in which new structures are introduced by axiomatisation and not necessarily motivated by informal ancestors. The final section discusses the consequences of string theorists' claim to finality for the theory's mathematical make-up. I argue that ontological reductionism as advocated by particle physicists and the quest for mathematically deeper axioms do not necessarily lead to identical results.

Key words: Jaffe-Quinn debate; rigour in string theory; final theories; Lakatos's philosophy of mathematics, John von Neumann; axiomatic method; theoretical mathematics; mathematical ontology.

In discussing the intimate relationship between theoretical physics and mathematics, scientists and philosophers alike keep wondering about what Eugene P. Wigner once called "the unreasonable effectiveness of mathematics in the natural sciences." (1960) Only few scientists find this effectiveness simply "reasonable" (Tisza, 1997); logical empiricism and the rigorous analytic-synthetic distinction have got out of fashion; mathematical platonists in Gödel's wake treat mathematics as a peculiar kind of empirical

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 197–222.

science (Köhler, 2002); realists and naturalists instead ponder whether mathematics is indispensable for the empirical sciences (Maddy, 1997; Colyvan, 2001; Leng, 2001). The light of mathematization, at any rate, does not spread homogeneously across the physical sciences; at places mathematics exhibits an “unreasonable uncooperativeness” (Wilson, 2001a) with scientists’ ontological needs.

These contributions document an increasing interest into a philosophy of mathematics that is not narrowly foundationalist in spirit but oriented at mathematical practice. (See the programmatic Corfield, 2003) The present paper intends to address a specific philosophical problem at the crossroads of theoretical physics and mathematics which in recent years has sparked controversies among scientists, but has not yet received much attention among philosophers of science – the new trend notwithstanding.

Wigner cited the examples of planetary motion, quantum mechanics, and quantum electrodynamics to show “that the mathematical language has more to commend it than being the only language which we can speak; … it is, in a very real sense, the correct language.” (Wigner, 1960, *p. 8*) What Wigner took as “the empirical law of epistemology,” (Ibid., p. 10) however, at bottom remained an act of faith. “The miracle of the appropriateness of the language of mathematics for the formulation of the laws of nature is a wonderful gift which we neither understand nor deserve.” (Ibid., p. 14)

Around the year 1990, or so it seems, another miracle, comparable in size but opposite in kind, has occurred in the domain of pure mathematics. It unveiled a converse of Wigner’s dictum. “Not only is mathematics the language of physics, but … in quite a large area of mathematical research today, *theoretical physics* has become the language of *mathematics. …* [W]e are confronted in mathematics with the difficulty of understanding the ‘unreasonable effectiveness of theoretical physics in mathematics’.” (Jaffe, 1997, *p. 138*)

In 1990, the theoretical physicist Edward Witten was awarded the Fields Medal for his contributions to geometry, which were largely stimulated by string theory. This prompted discussions among mathematicians and mathematical physicists as to how Witten’s mostly conjectural results ought to be appraised. No mathematician doubted that his representation of the Jones invariants of knots using Chern-Simons field theory was a major breakthrough that connected two hitherto unrelated subjects. Moreover, considerable parts of the results were quickly proven by pure geometers. But some conjectures were disproved and significant gaps remained. Had mathematicians witnessed “one of the most refreshing events in the mathematics of the 20th century” (Atiyah et al., 1994, p. 179), as Michael Atiyah felt, or did the dangers for mathematical rigour prevail unless one

prevented an uncritical copying of the methods at work in this singular success story?

In their 1993 article "'Theoretical Mathematics': Toward a Cultural Synthesis of Mathematics and Theoretical Physics", the mathematical physicists Arthur Jaffe and Frank Quinn proposed a set of prescriptions for the interaction between mathematicians and theoretical physicists that should foster mathematicians' receptivity of ideas from physics by safeguarding mathematical rigour against *uncontrolled* speculation. All authors of mathematical papers "should make a choice: either they provide complete proofs, or they should agree that their work is incomplete [conjectural] and the essential credit will be shared. Referees and editors should enforce this distinction, and it should be included in the education of students." (Jaffe/Quinn 1993, p. 10) Mathematics, so they argued, is either rigorous or theoretical. While rigorous results are final, even well-founded theoretical claims require corroboration by proof. The reliability of the literature, one of the prerequisites of progress and education, should be secured by a standard nomenclature that unambiguously flags 'theoretical' results as 'conjectures' (instead of 'theorems') that 'predict' (instead of 'show'), etc.

The Jaffe-Quinn paper provoked a broad controversy that is documented in no less than 16 responses by leading mathematicians and the authors' summary rejoinder in the next volume of the *Bulletin of the American Mathematical Society.*[78] Further voices on this "culture clash" between mathematics and physics were cited in the *Scientific American*. (Horgan, 1993) On the 1994 International Congress of Mathematical Physics in Paris, the Jaffe-Quinn debate was incitement enough to organise a round table "Physics and Mathematics."[79] However, its convenor Joel Lebowitz gave it a somewhat different thrust and asked participants to sketch problems of mathematical physics that did or did not contribute in advancing human understanding of nature.[80] By this shift away from the issues of reliability and proof, Lebowitz largely followed William Thurston's (1994) contribution to the debate. As regards the community of philosophers, the matter was taken up in the May 1997 issue of *Synthese*, but only mathematicians took a stand on "Proof and Progress in Mathematics."[81] Meanwhile, however, philosophers have devoted some thoughts to the Jaffe-Quinn debate as an example for problems of rigour and mathematical

[78] (Atiyah et al., 1994; Thurston, 1994; Jaffe and Quinn, 1994).

[79] See the Foreword of Daniel Iagolnitzer in (Iagolnitzer, 1995, p. 692).

[80] See Lebowitz's letter to the panellists in the materials distributed at the conference, p. 3; Atiyah's response on p. 4 expresses a certain discomfort with the change of focus.

[81] The only paper penned by a philosopher (Jaakko Hintikka's) is dedicated to the foundations of mathematics and, consequently, does not mention the Jaffe-Quinn debate.

ontology (Corfield, 1997, 2003; Aberdein, 2003; Davey 2003; Stöltzner, 2002a). A more profound discussion, however, is still lacking.

The present paper takes a first step in this direction by studying the Jaffe-Quinn debate against the backdrop of two philosophical accounts that are highly appreciative towards informal mathematical growth and conjectural results. After rehearsing the essentials of the debate (Section 1), I shall try to make sense of 'theoretical mathematics' within a Lakatosian approach. Particular emphasis will be given to the quasi-empirical character of mathematics and the dialectics of proofs and refutations. (Section 2) It is well-known that Lakatos's falsificationist methodology performs badly when it comes to modern axiomatized mathematics or to mathematical concepts without easily discernible informal ancestors. John von Neumann's conception of opportunistic axiomatics promises remedy by emphasizing the flexibility and the pragmatic virtues of axiom systems. (Section 3) An important element of the axiomatic method is the search for more general theorems or mathematically 'deeper' concepts that make the mathematical structure of the theory more conspicuous. Yet, "deepening the foundations" mathematically might yield concepts and entities that starkly differ from those favoured by physical reductionists. This problem is particularly pressing for string theory which, owing to the dim prospects of ever attaining experimental corroboration, can derive independent support only from the quality of its basic mathematical structures. Since string theorists, nonetheless, insist that their theory embraces all physical interactions and, accordingly, "has provided our first plausible candidate for a final theory," (Weinberg 1993, p. 169) the issue of mathematical rigour might influence the ontology of string theory. What if the basic physical concepts of the alleged 'Theory of Everything' are mathematically ill-founded? Moreover, and in stark contrast to these high aspirations, all attempts to squeeze at least some empirical predictions out of string theory by deriving, or at least establishing consistency with, those lower-level theories which have empirical support, typically involve perturbative methods. (Section 4)

It is true, a substantial part of the Jaffe-Quinn debate concerned the socio-cultural problems at the border between mathematics and theoretical physics. Who is to be credited, if a non-rigorous result is rigorously proven afterwards? Shall the community standards be safeguarded by explicit rules of conduct? Who controls whether these standards are obeyed? No doubt, these questions suggest interesting methodological considerations. But the philosophical core of the debate, to my mind, concerns ontological matters. If there is – all rules of conduct observed – a meaningful way of pursuing 'theoretical mathematics', what is it all about? If the concepts of 'theoretical mathematics' are analogous to those of theoretical physics, they must exceed

the stage of merely provisional speculations and become amenable to realistic or anti-realistic interpretations.

1. THE JAFFE-QUINN THESES

The main motivation for Jaffe and Quinn to stipulate a clear-cut distinction between mathematical speculation and rigorous proofs derived from their philosophical understanding of the discipline and its history. "Modern mathematics is nearly characterized by the use of rigorous proofs. This practice, the result of literally thousands of years of refinement, has brought to mathematics a clarity and reliability unmatched by any other science." (Jaffe and Quinn, 1993, p. 1) They distinguish two stages of mathematical research.

> First, intuitive insights are developed, conjectures are made, and speculative outlines of justifications are suggested. Then the conjectures and speculations are corrected; they are made reliable by proving them. We use the term *theoretical* mathematics for the speculative and intuitive work; we refer to the proof-oriented phase as *rigorous* mathematics. (Ibid.)

This terminology expresses a functional analogy between rigorous proof and experimental physics. Both correct, refine and validate the claims of their theoretical counterparts.

Proofs serve two main purposes. First, they "provide a way to ensure the reliability of mathematical claims." (Ibid., p. 2) "Second, the act of finding a proof often yields, as a by-product, new insights and unexpected new data." (Ibid.) Hence, 'theoretical mathematics' is built on an asymmetry of proof and conjecture. Posing a conjecture does not necessarily involve proof; heuristics is subordinated to the justificatory role of proof.

1.1 The Theoretical and the Experimental

Jaffe and Quinn are aware that the analogy between physics and mathematics is limited: "we are not suggesting that proofs should be called 'experimental' mathematics. There is already a well-established and appropriate use of that term, namely to refer to numerical simulations as tests of mathematical concepts." (Ibid., p. 2) Still, many mathematicians dispute the reliability of computer proofs. Armand Borel criticises where the functional distinction is situated. "Roughly, the experimental side is the investigation of special cases … and the theoretical side is the search of

general theorems. In both, I expect proofs of course, and I categorically reject a division into two parts, one with proof, the other without." (Atiyah et al., 1994, p. 180) According to Saunders Mac Lane, the comparison of proofs with experiments is faulty. "Experiments may check up on a theory, but they may not be final; they depend on instrumentation, and they may even be fudged. The proof that there are infinitely many primes … is always there. … Mathematics rests on proof – and proof is eternal." (Ibid., p. 193)

Morris W. Hirsch emphasises that "the nonrigorous use of mathematics by scientists, engineers, applied mathematicians and others is in fact more complex than simple speculation." (Ibid., p. 186) It involves the use of mathematical language for 'narrative purposes', in particular if the result has already been experimentally verified. Karen Uhlenbeck holds that "'theoretical mathematics' already exists. It is called 'applied mathematics', a much bigger field than pure mathematics. … Only the combined elitism of very pure mathematics and high-energy fundamental physics would claim that its own brand of speculative and applicable mathematics should have a special name." (Ibid., p. 202) What about non-linear dynamics or mathematical biology? A representative of this alleged elite, the string theorist Albert Schwarz, equally considers the terminology inappropriate as a common name for heuristic mathematics and theoretical physics. But he stresses the peculiarity of string theory within applied mathematics: Today, theorists "are not able to extract reliable predictions from string theory because this is connected with enormous mathematical difficulties. The physicists have chosen the only possible way: to analyze carefully the mathematical structure of string theory." (Ibid., p. 197) This is exactly Jaffe and Quinn's point: Theoretical physicists "have found a new 'experimental community': mathematicians … who provide them with reliable new information about the structure they study." (Jaffe and Quinn, 1993, p. 3) Indeed quite a singular situation if measured against the whole of the interactions between physicists and mathematicians.

1.2 The Ontology of Theoretical Mathematics

'Theoretical mathematics' tacitly requires some sort of 'quasi-empirical' ontology. "For if we don't assume that mathematical speculations are about 'reality' then the analogy with physics is greatly weakened – and there is no reason to suggest that a speculative mathematical argument is a theory of anything, any more than a poem or novel is 'theoretical'" (Atiyah et al., 1994, p. 186) writes Hirsch. Mac Lane, in contrast, advocates austerity: "If a result has not yet been given valid proof, it isn't yet mathematics: we should strive to make it such." (Mac Lane, 1997, p. 151) To his mind, all other assertions root in the misconception of "set theory as THE foundation of

mathematics, and so sometimes [philosophers of mathematics] eagerly spread the gospel that mathematics is the study of an ideal realm of sets – set theoretic platonism." (Ibid.)

Also some of those advocating 'theoretical mathematics' without the Jaffe-Quinn strictures count on foundational support. René Thome uses, expectedly, Gödel's incompleteness theorem to bolster his claim "that rigor can be no more than a local and sociological criterion." (Atiyah et al., 1994, p. 203) But Mac Lane (1997) is fully right to stress that Gödel's theorem concerns very specific systems, those that admit a Gödel numbering.

It appears to me, however, that foundationalist aspects in the narrow sense are of only minor importance to the Jaffe-Quinn debate. What Mac Lane, in effect, rejects is string theoreticians' belief that the mathematical structures they have heuristically justified necessarily exist in some sense or other, the more concrete determination of which is generously left to rigorous mathematicians. Citing historical examples from classical applied mathematics, Mark Wilson has adequately baptized such a stand as "lazy mathematical optimism". It is characterised by the belief that "*every real-life physical structure can* [a priori] *be expected to possess a suitable direct representative within the world of mathematics.*" (Wilson, 2000, p. 297) In view of the many mismatches between mathematics and physics, even in domains as profane as continuum mechanics or elasticity theory, lazy optimism becomes untenable. But there is still an honest version of mathematical optimism which, on Wilson's account, goes back to Leonhard Euler.[82] After diagnosing the lack of a suitable mathematical structure to treat a physical problem, the mathematical optimist can try to liberalize mathematical ontology so as to include all physically possible solutions or devise new concepts that have been ill-defined within the previously accepted mathematical framework. The difference between lazy and honest optimism lies precisely in whether one succeeds in providing such a liberalized ontology or concepts that are well-defined in a suitable sense. 'Theoretical mathematics' is the first step of the honest optimist, and the Jaffe-Quinn debate concerns precisely the question at which point we can trust in mathematical honesty.

Notice that in applied mathematics matters stand better than in string theory because one typically possesses independent physical evidence that a system described by the mathematical equation under scrutiny exists as a well-entrenched entity. Take, for instance, the turbulences of a water flow through a pipe or a fracturing rod. In contrast, all available evidence for string theory obtains only by way of other physical theories which have

[82] But also many of Hilbert's problems were like this; see (Stöltzner, 2004).

mathematical problems of their own. Moreover, the mathematical transitions to these theories involve perturbative methods and unpredictable symmetry breaking mechanisms. All problems of experimental verification notwithstanding, Witten aspires beyond mathematical optimism: "when a mathematical result is really relevant to a physics problem it often happens that, turning things around, the result can be deduced from the behaviour of the physics problem." (Iagolnitzer, 1995, p. 704) Such claims make the problems of mathematical ontology even more pressing.

1.3 Linearising Mathematical Progress

A considerable part of Jaffe and Quinn's paper and the responses dealt with the controversial lessons historical examples teach. Aside from undisputed success stories, there are also 'cautionary tales' that demonstrate that relaxing standards and relying on intuitions occasionally was a hindrance – or even disastrous – for a budding research programme.

The ideal attitude in the contact between mathematics and physics was assumed by mathematical physicists, such as "D. Hilbert, F. Klein, H. Poincaré, M. Born, and later H. Weyl, J. von Neumann, E.P. Wigner, M. Kac, A.S. Wightman, R. Jost, and R. Haag. … These people often worked on questions motivated by physics, but they retained the traditions and values of mathematics," (Jaffe and Quinn, 1993, p. 4) to wit: rigour, scholarship, and knowledge of the literature. Their speculations, on the other hand, were addressed to physicists.

Jaffe and Quinn discuss essentially three types of success stories. (i) Brilliant conjectures have inspired the development of whole fields. "The Hilbert problem list, of amazing breadth and depth, has been very influential in the development of mathematics in this century." (Jaffe and Quinn, 1993, p. 6) (ii) Conjectures that were accompanied by technical details or even an outline of a proof, such as the Weil conjectures, have initiated entire research programmes. (iii) Famous conjectures, among them Fermat's Last Theorem, can be highly motivating if they turn out to be the corollary of a general theorem.

"Most of the experiences with theoretical mathematics have been less positive." Here are two examples: At the beginning of this century, the 'Italian school' of algebraic geometry "collapsed after a generation of brilliant speculation. … In 1946 the subject was still regarded with such suspicion that Weil felt he had to defend his interest in it." The historian Jeremy J. Gray, however, emphasised that the Italian school "by modern standards … seems to lack rigour – but this perception is modern, and due to Zariski, who also brought new questions to bear (such as arbitrary fields)." (Atiyah et al., 1994, p. 185)

The way how Jaffe and Quinn classify success and failure suggests that they prefer a linear growth that is initiated by a conjecture or a theoretical result from which a research programme, though slowly and piecemeal, immediately takes off and yields rigorous results. Ruptures should best be avoided, because they hamper progress in the long run, although momentary growth might be considerable. Lasting historical progress requires scholarship and a reliable literature. Moreover, Jaffe and Quinn expect scientific journals to provide the same reliability as textbooks. Put in a nutshell: although the authors do not write a piece of Whig historiography, they seem to wish that the growth of mathematics be guided to conform to a Whiggish account.

In his Response, Atiyah lodges a protest against the linear growth model:

> [Jaffe and Quinn] present a sanitized view of mathematics which condemns the subject to an arthritic old age. They see an inexorable increase in standards and are embarrassed by earlier periods of sloppy reasoning. But if mathematics is to rejuvenate itself and break new ground it will have to allow for the exploration of new ideas and techniques which, in their creative phase, are likely to be dubious as in some of the great eras of the past. Perhaps we now have high standards of proof to aim at but, in the early stages of new developments, we must be prepared to act in more buccaneering style. (Ibid., p. 178.)

"However, a buccaneer is a pirate, and a pirate is often engaged in stealing," retorts Mac Lane: "Buccaneers have no place in mathematics." (1997, p. 150) And thus, so one might continue, the Jaffe-Quinn prescriptions represent much-needed maritime law. Yet, a bad metaphor does not make a faulty argument. To my mind, Atiyah is plainly right to argue that the linear model of mathematical growth is incapable of assessing the boost of knowledge that geometry has received from string theory.

2. A LAKATOSIAN PERSPECTIVE

In his much-read dialogue *Proofs and Refutations,* Imre Lakatos outlined a philosophical methodology in which the growth of mathematical knowledge is driven by the dialectics of proofs and refutations and by the continuous interaction of heuristics and validation. Lakatos's rational reconstruction of the history of the Euler conjecture thus challenges the linear account of history to the same extent as his methodology attributes due space to 'theoretical mathematics'.

In this section I argue that (i) any hermetic separation between heuristics ('theoretical mathematics') and proof (rigorous, experiment-like

mathematics) ignores the subtle dialectic between conjectures and proofs. (ii) In the history of science, we are not faced with a one-shot interaction conjecture-proof, but with the course of a mathematical research programme. (iii) A Lakatosian perspective gives ontological justification to 'theoretical mathematics' because it provides a mechanism to create the ontology appropriate for a certain mathematical quasi-fact. (iv) But due to his unrestrained fallibilism, Lakatos failed to appraise the virtues of axiomatization. Moreover, it is unclear what stuff mathematical research programmes are about.

2.1 The Mechanism of Proofs and Refutations

Jaffe-Quinn and Lakatos share a common starting point: 'Proof' stands "for a thought-experiment – or 'quasi-experiment' – which suggests a decomposition of the original conjecture into subconjectures or lemmas, thus embedding it in a possibly quite distant body of knowledge." (Lakatos, 1976, p. 9) Thought-experiments follow an initial naive trial and error phase. In the example around which Proofs and Refutations is built, they correspond to a stretching and a triangulation of a polyhedron. The lemmas, or subconjectures, ensure that the 'thought-experimental' steps of the proof are permissible, such that the proof can validate the conjecture. But, quite similar to the development of experimental techniques in science, the decomposition is informative even if validation does not obtain. And if it does, its main achievement is to provide an improved basis for scepticism.

Refutations are suggested by counterexamples that either concern the conjecture (*global* counterexamples) or the lemmas (*local* counterexamples). (i) Global, but not local counterexamples logically refute the conjecture. They are what most mathematicians would call a counterexample. (ii) If a global counterexample is also local, it does not refute the theorem, but confirms it. (iii) Local, but not global counterexamples show a weakness of the theorem, such that one has to search for modified lemmas. Cases (ii) and (iii) are not genuinely logical, but *heuristic* counterexamples.

The imaginary class of *Proofs and Refutations* discusses various strategies to handle counterexamples. *Monster-barring,* rejects the global counterexample as "a pathological case" (Ibid., p. 14) of a polyhedron by modifying the latter's definition in a suitable way. It decreases the domain of validity of the conjecture, only. But counterexamples abound despite such linguistic ad hoc remedies. Theoretical physicists often apply this strategy by suggesting a natural 'physical definition' of the mathematical concept in question or, more operationally, by rejecting 'pathologies' as beyond the scope of the studied model. The history of theoretical physics, however,

teaches us that monsters may always return because many mathematical structures eventually have turned out to be of physical interest.

The next strategy, *exception-barring,* restricts the domains of both the conjecture and of the guilty lemma. But the strategic withdrawal could have been too radical, and the method still does not exploit the proof. *Monster-adjustment*, the third strategy, also concerns the domain of validity of the basic concepts. "Monsters don't exist, only monstrous interpretations." (Ibid., p. 31) This therapeutic method strives to see an example in the alleged counterexample by finding a suitable interpretation. In a footnote, Lakatos rejects it. "Nothing is more characteristic of dogmatist epistemology than its theory of error." (Ibid., p. 31, fn. 3) Later, however, he noted that monster-adjustment could be empirically progressive in science.[83] Indeed, empirical science might ease monster-adjustment by singling out 'physical' or 'realistic' interpretations. But this cannot compensate for faulty proofs and mathematically ill-defined concepts.

In Cauchy's days, mathematics matured through an appropriate estimation of proof-analysis. The method of *lemma-incorporation* "upholds the proof but reduces the domain of the main conjecture to the very domain of the guilty lemma." (Lakatos, 1976, p. 34) In this way, the lemma refuted by the counterexample is built into the conjecture. Hence, proofs improve a conjecture, even if they do not prove it. This "displays the fundamental dialectical unity of proof and refutations." (Ibid., p. 37) But on principle, one has to incorporate expectable, but not yet known, counterexamples. This reveals that lemma incorporation proceeds through constant overstatements, by attempting to keep as much as possible from the initial thought-experiment and its heuristics. While this method emphasises the heuristic role of proof, exception-barring focuses on validity and advances through a series of understatements. Accordingly, it corresponds to the 'better-safe-than-sorry' strategy prescribed by Jaffe and Quinn. Lakatos stresses that a careful proof-analysis, which constantly suspends unnecessary restrictions, makes the method of proof and refutations "a limiting case of the exception-barring method." (Ibid.)

In a later footnote, Lakatos recalls his "deliberate *mixed* usage of the justificationist term 'proof' and of the heuristic term 'proof'." (Lakatos, 1978b, p. 135, fn. 3) More than being a mere by-product, as Jaffe and Quinn hold, heuristics stands on a par with rigorous justification. Exploiting the heuristic possibilities of the proof might overthrow the initial conjecture and supplant it by newly formulated theorems. Moreover, "different proofs

[83] See fn. 3 on p. 63 of "Falsification and the methodology of scientific research programmes" in (Lakatos, 1978a).

[better: 'improofs'] of the same naive conjecture lead to quite different theorems." (Lakatos, 1976, p. 65) This leads to the full-blown method of proof*s* and refutations. The important point for the Jaffe-Quinn debate is that Lakatos's methodology of mathematical growth does not play down the role of proofs. Emphasising their heuristic role, on the contrary, puts proof thought-experiments into the core of historical development. The 'speculations' that make up 'theoretical mathematics' only concern single conjectures – be they supplemented with a proof-technique or not.

Later, Lakatos considered it to be a leitmotiv of *Proofs and Refutations* that "*one may bravely – and profitably – go on to 'explain' a hypothesis known to be false.*" (Lakatos, 1978b, p. 176 fn. 3) One may not even need a conjecture to start proving or testing by means of thought-experiments. Being ready to give up the naive conjecture, a more general theorem might be easier to prove. An extended version of the method of analysis-synthesis driven by heuristic and validating thought-experiments contains possible *occult hypotheses*. For the evaluation of certain real integrals, the introduction of complex numbers is a necessary prerequisite. For geometry, string theory has provided brilliant occult hypotheses and new concepts that were generated by Witten's heuristic proofs. Hence, from a Lakatosian perspective, 'theoretical mathematics' is not an intermediate step in the sequence from conjecture to proof, but rather an indispensable element of the extended analysis-synthesis circuit.

Over the course of history, concepts grow – sometimes even by rather wild extensions, such as the Peano curve which fills a two-dimensional surface. Lakatos argues that it was not the monster barrers who contracted concepts, but the refutationists expanded them to cover objects unintended by the naive conjecture. "*Often, as soon as concept-stretching refutes a proposition, the refuted proposition seems such an elementary mistake that one cannot imagine that great mathematicians could have made it.*" (Lakatos, 1976, p. 87, fn. 1) But such an accusation, which also characterises Jaffe and Quinn's 'cautionary tales', neglects precisely that sort of concept growth that reaches beyond a mere change in rigour. In this respect, most negative Responses to "Theoretical Mathematics" still have too narrow a focus.

Lakatos severely criticises the widespread view that an 'informal' proof is a formal proof with gaps. "[T]o suggest that an informal proof is just an incomplete formal proof seems to me to be to make the same mistake as early educationalists did, when, assuming that a child was merely a miniature grown-up, they neglected the direct study of child-behaviour." (Lakatos, 1978b, p. 63) In the same vein, Jaffe and Quinn distinguished two groups of informal reasoning according to whether the conjecture in question could subsequently be proven. However, justificationists unduly maximise

historical continuity by separating the hard formal kernel that holds true still today from the erroneous 'metaphysical' interpretation. Thus, Poincaré's results are still not fully assessed if one insists (against Jaffe-Quinn) that they matched the standards of the day.

Instead, heuristic power decides the faith of a research programme that was ventured from an initial conjecture. Formal arguments alone, however, cannot make up the core of a research programme without taking into account 'metaphysical' heuristics. The latter might even be linked to the heuristic support derived from physics. This was the case for Dirac's delta function which had become part and parcel of quantum theory long before it was made mathematically rigorous. While von Neumann (1932) opposed it, Laurent Schwartz relaxed the ontology of the concept of function by defining generalised functions (distributions), thus winning honesty for Dirac's mathematical optimism.

2.2 Quasi-empirical Ontology

Let me now show how it is possible to obtain some quasi-ontological backing for 'theoretical mathematics' from Lakatos's criticism of foundationalism. While *Euclidean* theories are built on indubitable axioms from which truth flows down through valid inferences, in quasi-empirical theories truth is injected at the bottom by virtue of a set of accepted basic statements. In the latter case, truth does not flow downward from the axioms, but falsity is retransmitted upward. "[I]n a quasi-empirical theory the (true) basic statements are *explained* by the rest of the system.“ (Lakatos, 1978b, p. 28f.) And it is only the flow of truth that is at stake; "*a theory which is quasi-empirical in my sense may be either empirical or non-empirical in the usual sense.*" (Ibid., p. 29) Theoretical physics is, of course, quasi-empirical and empirical. Among the basic statements of a conceptually mature theory which 'explain' physical facts, genuinely mathematical ones can be found alongside basic empirical facts, such as measurable constants of nature. But can one, quite generally, consider 'theoretical mathematics' as quasi-empirical without counting on such 'empirical axioms'?

In Lakatos's view, the borderline between mathematics and the sciences is drawn by the mode of verification: "If mathematics and science are both quasi-empirical, the crucial difference between them, if any, must be in the nature of their 'basic statements' or 'potential falsifiers'." (Ibid., p. 35) Contradictions are the typical *logical* falsifiers.

> But if we insist that a formal theory should be the formalization of some informal theory, then a formal theory may be said to be 'refuted' if one of its theorems is negated by the corresponding theorem of the informal

> theory. One could call such an informal theorem a *heuristic falsifier* of the formal theory. Not all formal theories are in equal danger of heuristic refutation in a given period. For instance, *elementary group theory* is scarcely in any danger: in this case the original informal theories have been so radically replaced by the axiomatic theory that heuristic refutations seem to be inconceivable. (Ibid., p. 36)

On the other hand, after the destruction of naive set theory by logical falsifiers, one cannot speak any longer of set-theoretical facts. Nevertheless, one might still continue to consider it to be the unifying basis of mathematics. Hence, the question of mathematical facts rests upon a subtle interaction between the informal and the formal level. For the Jaffe-Quinn debate, this entails that those objects of informal 'theoretical mathematics', which are blatantly inconsistent, can hardly count as quasi-empirical mathematical facts in Lakatos's sense. Moreover, attitudes as relaxed concerning rigour as Thom's cannot count on Lakatos because they neglect the heuristic power or rigour.

2.3 On Mathematical Research Programmes

At the time of his death, Lakatos had planned to apply the methodology of scientific research programmes (MSRP) to the history of mathematics. A footnote in the 1970 paper launching MSRP reads as follows: "Unfortunately in 1963-4 I had not yet made a clear terminological distinction between theories and research programmes, and this impaired my exposition of a research programme in informal, quasi-empirical mathematics." (Lakatos, 1978a, p. 52, fn. 1)

A research programme is defined by its hard core, which is tenaciously defended by negative heuristics. It is surrounded by a protective belt of quite flexible positive heuristics, which constantly put forward auxiliary hypotheses against anomalies. The programme supplies a conceptual framework and contains a powerful problem solving machinery. As there is no sharp distinction between theory and experiment, rival theories do not encounter each other one-by-one; rather, sequences of theories within a research programme compete with rivals in the face of a larger body of empirical evidence. This makes it possible to establish internal criteria of progress. A programme is progressive if each theory has excess empirical content over its predecessors, and if some of the predicted novel facts are corroborated. A programme is degenerating if its theories are only fabricated to accommodate known facts by way of a (content-decreasing) linguistic reinterpretation.

So far, all this fits quite neatly to Lakatos's philosophy of mathematics, which emphasises the inseparability of conjecture and proof. Lakatos's concept of growth elucidates why string theory has to seek the vicinity of mathematics to get its excess content corroborated – at least as a mathematical quasi-fact. MSRP seems to be a reasonable approach, if theories abound. If, in contrast, theories grow more slowly than empirical facts are provided, Lakatos could call hardly any programme progressive. To Lakatos, unifications that are usually held in high regard by mathematicians cannot count as progressive problemshifts, unless they lead to concepts of results unknown so far, such as the five exceptional Lie groups.

The ambiguity of the notion of progress is mitigated by a dose of lenience. "Criticism is not a Popperian quick kill, by refutation. Criticism is always constructive: there is no refutation without a better theory." (Ibid., p. 6) Content-decreasing strategies can be temporarily employed, if anomalies abound and technical difficulties slow down possible predictions. Then one does not accept anomalies as genuine counterexamples, and one allows for a certain autonomy of theory. "*Mature science consists of research programmes in which not only novel facts but, in an important sense, also novel auxiliary theories, are anticipated; mature science – unlike pedestrian trial-and-error – has 'heuristic power'* ... [which] generates *the autonomy of theoretical science.*" (Ibid., p. 88) Hence, MSRP justifies the autonomy of 'theoretical mathematics' – even if it is not empirically progressive due to incomplete proofs – but rejects its neat separability from rigorous mathematics.

It seems then that the general scheme of progress versus degeneration can be easily translated from empirical science to mathematics. A programme is theoretically progressive if it proposes fruitful concepts and techniques; it progresses empirically if it solves interesting problems (particularly those posed in another field).

What stuff a methodology of mathematical research programmes could be all about, seems less clear. In this respect, Lakatos's quasi-empirical ontology needs qualification. David Corfield argues that "rivalry between research programmes concerns high level issues." (1998, p. 276) These levels come about because, in comparison to physical science, "[m]athematics appears to have an extra degree of freedom at this [basic] level [where battles are usually fought out] which makes it improbable that programmes will be in direct competition for precisely the same territory." (Ibid., p. 295) Hard cores do not simply boil down to axioms, and there are no universally agreed-upon facts – as it was the case in the paradigmatic competition between ondulatory and emission theory in 17th and 18th century optics. Beliefs or general aims, might enter the hard core, shifting emphasis away from conjectures as the sole driving force of research programmes. In

fact, conjectures might "be decided one way or the other by an uninformative proof or an uninstructive counterexample." (Ibid., p. 280) Introducing higher-level issues "would bring the hard core and positive heuristic closer, thereby threatening to collapse the whole construction." (Ibid., p. 281) But this move seems necessary, if one wants to assess the phenomenon – quite common in mathematics – that two theories emerge from a single common problem, or converge into one, although they do not dispute the same area of quasi-empirical facts.

3. JOHN VON NEUMANN'S OPPORTUNISTIC AXIOMATICS

When Lakatos diagnosed "A renaissance of empiricism in the recent philosophy of mathematics", John von Neumann was among his witnesses. Unfortunately, he only assessed the foundationalist themes in "The Mathematician", but not the relationship between their views about the mathematical method. This surprising neglect might be a consequence of Lakatos's far-reaching aversion against the axiomatisation of science that was rooted in his stubborn insistence that no proof whatsoever be considered final – not even relative to the (always revocable) acceptance of certain axioms and a suitable metatheory.[84]

Identifying axiomatization with absolute finality blatantly misrepresents von Neumann's methodological stance. More than Hilbert, von Neumann gave the axiomatic method a decidedly pragmatic twist that allows one, so the present section argues, to avoid several shortcomings of the Lakatosian account. (i) Although von Neumann regards proofs as more definitive than Lakatos, mathematical rigour is not immutable and the reliability of mathematics, accordingly, comes close to that of well-established scientific facts. (ii) What is more, there exists no neat separation between the theoretical branches of the empirical sciences and mathematics. (iii) Rigorous axiomatization, on the other hand, proves fertile even in case the basic concepts of a science are not yet clarified and empirical evidence is poor because mathematization permits great flexibility and opportunism in concept formation. (iv) Quite in line with Lakatosian methodology, mathematics is itself capable of heuristic development relevant to the sciences because its best inspirations stem from empirical problems. (v) Mathematics and the sciences share some aesthetic criteria of success. Mathematics proper disposes of a further aesthetic criterion that concerns the

[84] Cf. Worrall and Zahar's editors' note on p. 138 of (Lakatos, 1976).

maturity of theories and the architecture of proofs. They are good candidates for the higher level issues mathematical research programmes compete for.

3.1 Rigour and the Role of Physics for Mathematics

In *The Mathematician*, von Neumann remembers "how humiliatingly easily my own views regarding the absolute mathematical truth changed … three times in succession." (Neumann 1947, p. 6)

> The main hope for justification of classical mathematics – in the sense of Hilbert or of Brouwer and Weyl – being gone, most mathematicians decided to use that system anyway. After all, classical mathematics … stood on at least as sound a foundation as, for example, the existence of the electron. Hence, if one is willing to accept the sciences, one might as well accept the classical system of mathematics. (Ibid., p. 6).

But the erosion of meta-theory does not lead to the complete demise of mathematical rigour. Although any particular set of basic propositions can be doubted, mathematics "establishes certain standards of objectivity, certain standards of truth … rather independently of everything else." (Neumann, 1954, p. 478). This objectivity does not contradict the historical fact that many non-rigorous arguments were accepted – either with a certain sense of guilt or due to *bona fide* disagreements as to whether a particular proof was really a proof. Rather do the historical fluctuations of rigour teach a lesson that is of great importance to ontology of 'theoretical mathematics.'

"The variability of the concept of rigor shows that something else besides mathematical abstraction must enter into the makeup of mathematics" (Neumann, 1947, p. 4). Here the empirical sciences are called upon. "The most vitally characteristic fact about mathematics is … its quite peculiar relationship … to any science which interprets experience on a higher than purely descriptive level." (Ibid., p. 1) This relationship has two sides: On the one side,

> [i]n modern empirical sciences it has become a major criterion of success whether they have become accessible to the mathematical method or to the near-mathematical methods of physics. Indeed, throughout the natural sciences an unbroken chain of pseudomorphoses, all of them pressing toward mathematics, and almost identified with the idea of scientific progress, has become more and more evident. (Ibid., p. 2)

On the other side, "[s]ome of the best inspirations of modern mathematics (I believe, the best ones) clearly originated in the natural sciences."(Ibid.) Von Neumann provides two examples.

(i) The origin of *geometry* in antiquity was empirical; "it began as a discipline not unlike theoretical physics today." (Ibid.) Euclid's ensuing postulational treatment even served as a model for Newton's *Principia.* The 'de-empirisation' of Euclidean geometry was never quite completed until with Hilbert the axiomatic method itself obtained a new abstract meaning and was extended to non-Euclidean geometries. Yet, in the form of general relativity empiry has not only the final say, but also initial doubt stems from there: "The prime reason, why, of all Euclid's postulates, the fifth was questioned, was clearly the unempirical character of the concept of the entire infinite plane which intervenes there, and there only." (Ibid., p. 3)

(ii) *Calculus*, Newton's fluxions in particular, was explicitly created for the purpose of celestial mechanics. "An inexact, semiphysical formulation was the only one available for over a hundred and fifty years after Newton!" (Ibid.) Despite major advances, "[t]he development was as confused and ambiguous as can be. … And even after the reign of rigor was essentially re-established with Cauchy, a very peculiar relapse into semiphysical methods took place with Riemann." (Ibid., p. 3f.)

Hence, quite generically, those scientific theories which cannot avail themselves of previously created mathematical structures are likely to incite their own mathematics that sets out in a rather informal way. Thus a certain part of empirical science as a whole becomes the informal ancestor of a mathematical discipline. While these examples could be subsumed under the Lakatosian outlook, von Neumann was well aware that there exist counterexamples. "There are various important parts of modern mathematics in which the empirical origin is untraceable" (Ibid., p. 6) or very remote, such as topology or abstract algebra. "Two strange examples are given by differential geometry and by group theory: they were certainly conceived as abstract, nonapplied disciplines. … After a decade in one case, and a century in the other, they turned out to be very useful in physics. And they are still mostly performed in the indicated, abstract, nonapplied spirit." (Ibid., p. 7) Hence, there must be specific and self-contained mathematical criteria of success which, on the other hand, permit a rather smooth transition from empirical science to mathematics.

3.2 On Progress in the Science and Mathematics

To von Neumann, the prevailing attitude in science is opportunism: the sciences "mainly make models" (Neumann, 1955, p. 492) which are valid over limited scales only.

> The criterion of success of such a theory is simply whether it can, by a simple and elegant classifying and correlating scheme, cover very many

> phenomena, which without this scheme would seem complicated and heterogeneous, and whether this scheme covers phenomena which were not considered at the time when the scheme was evolved. (Neumann, 1947, p. 7)

"Simplicity is largely a matter of historical background … and it is very much a function of what is explained by it," (Neumann, 1955, p. 492) to wit, how heterogeneous the material covered by the explanation is. Accordingly, simplicity and unificatory power have to be equilibrated. In contrast to MSRP, von Neumann attributes little weight to whether prediction occurs before of after the fact. Heterogeneity ranks higher, in particular "confirmations in areas which were not in the mind of anyone who invented the theory." (Ibid., p. 493) Both simplicity and heterogeneity are "clearly to a great extent of an aesthetical nature." (Ibid.)

Mathematics proper possesses a further measure of progress. "One expects a mathematical theorem or a mathematical theory not only to describe and to classify in a simple and elegant way. … One also expects 'elegance' in its 'architectural', structural makeup," (Neumann, 1947, p. 9) e.g., a surprising twist in the argument which immediately makes a point very easy, or some general principle which explains why difficulties crop up and which reduces the apparent arbitrariness. "These criteria are clearly those of creative art" (Ibid.) so that

> the subject begins to live a particular life of its own and is better compared to a creative one, governed by almost entirely aesthetical motivations, than to anything else and in particular, to an empirical science. … As a mathematical discipline travels far from its empirical source … it is beset with very grave dangers. It becomes more and more purely aestheticizing, more and more purely *l'art pour l'art.* (Ibid., p. 9)

The field is then in danger of developing along the line of least resistance and might "separate into a multitude of insignificant branches." (Ibid., p. 9) "[W]henever this stage is reached, the only remedy seems … to be a rejuvenating return to the source: the reinjection of more or less directly empirical ideas." (Ibid., p. 9) Rejuvenation sounds less threatening to mathematicians than Atiyah's buccaneers. To von Neumann's mind, even in mathematical subdisciplines that possess a well-entrenched conception of rigour, such as geometry, a temporal return to 'theoretical mathematics' might be on the agenda.

While the aesthetic criteria of success bring mathematics and theoretical physics close to one another, von Neumann locates major differences regarding their actual *modus procedendi*. Even without signs of degeneration, mathematics is more finely subdivided because often the

selection of problems itself is aesthetically oriented. These divisions permit competition between mathematical research programmes as regards the above-mentioned higher-level issues. Theoretical physics, to the contrary, is typically highly focused to resolve an internal difficulty or to solve a problem that was posed by experimental results. Once a break-through is reached, "the predictive and unifying achievements usually come afterward." (Neumann, 1947, p. 8) "[T]he problems of theoretical physics are objectively given; and, while the criteria which govern the exploitation of a success are … mainly aesthetical, yet the portion of the problem, and that which I called above the original 'break-through', are hard, objective facts." (Ibid., p. 8)

Thus, the manifold methodological bridges between mathematics and physics do not rest upon a joint ontological domain. But, there exists a joint domain of Lakatosian quasi-ontology where both in mathematics and theoretical physics basic principles are explained by the theorems ensuing from them. Retransmission of falsity is not foreign to the axiomatic method. Already within Hilbert's conception, the axioms in first place had to be complete, that is, permit to derive all laws of the respective field, be they mathematical or physical in kind. Hilbert understood the axiomatic method as a critical companion to evolving scientific theories. Apart from checking completeness, the mathematician had to establish the internal and external consistency of the axiom system, and examine whether the axioms were mutually independent or whether they could be replaced by fewer, simpler or mathematically more natural axioms – a method which Hilbert called 'deepening the foundations.' (See Section 4.) Hilbert understood external consistency in a rather loose sense: the axioms should not contradict neighbouring domains of facts. Internal consistency, on the other hand, was established by construing appropriate number fields, thus playing internal consistency back to the consistency of arithmetic. Only this last step, the idea of establishing an absolute mathematical ontology, proved unfeasible after Gödel's Incompleteness Theorems. Completeness in Gödel's sense meant that all mathematical statements that are true within an axiom system are provable by internal operations; hence Gödel-completeness was a purely syntactic property. Hilbert's above-mentioned requirement of completeness, however, was semantic and remained unassailed if re-interpreted as a brand of 'post-formal mathematics'. With this term Lakatos denoted the classification of possible representations; for "axioms in the most important mathematical theories implicitly not just define one, but quite a family of structures." (Lakatos, 1978b, p. 69) Some of these models are intended, others are not, and among those are monsters that can easily be barred and monsters that cannot.

Dropping the requirement of internal consistency in the absolute sense but granting that mathematical concepts, relatively consistent with respect to arithmetic or set theory, were still more reliable than concepts of the empirical science von Neumann changed Euclideanist ontology into quasi-empirical ontology. Maintaining, however, that mathematical ontology was awarded by axiomatisation and giving the axiomatic method a pragmatic twist, he was simultaneously able to stress the parallels to the empirical sciences and avoid the problematic Lakatosian requirement that every formal theory must have an informal ancestor. This move gave heuristics a broader domain of application because some aesthetic and empirical inspirations are possible only after the respective field has been formalised to a sufficient degree.

It is true, axiomatisation implicitly contains the danger of becoming static, thus excluding the possibility of a richer theory. For this reason, Lakatos warned against early Euclideanisation because "we have no guarantee that our formal system contains the full empirical or quasi-empirical stuff in which we are really interested and with which we dealt in the informal theory. There is no formal criterion as to the correctness of formalization." (Lakatos, 1978b, p. 67). Agreed, probability theory without the Lebesque integral or algebra without complex numbers would be much poorer theories and lack key theorems. It is unclear whether these poorer theories would ever be diagnosed of aestheticism by internal mathematical criteria. Moreover, opportunism alone does not prompt mathematicians to seek rejunivation from the empirical sciences. This is, to my mind, one of the major reasons why 'theoretical mathematics' is indispensable for enriching the content of mathematics.

To sum up, the axiomatic method, as understood by Hilbert and von Neumann, represents a dynamical process in which the mathematical quality of an axiom system and the adequacy of the scientific theory derivable from it are constantly under scrutiny. This makes the study of a particular axiom system together with its rigorous and theoretical machinery a good candidate for the core of a mathematical research programme.

3.3 Mathematization as Theorizing

The opportunism of the axiomatic method, in von Neumann's understanding, is not only expressed in the central role of aesthetic criteria of success but it also derives from the great conceptual flexibility inherent in mathematics. Only this unleashes the Lakatosian conceptual dynamics between heuristics and rigour, or between proofs and refutations. It is also this flexibility due to which the honest mathematical optimist might always prevail in the long run.

> I feel that one of the most important contributions of mathematics to our thinking is, that it has demonstrated an enormous flexibility in the formation of concepts, a degree of flexibility to which it is very difficult to arrive in a non-mathematical mode. (Neumann, 1954, p. 482)

Two aspects of this flexibility are of specific importance. First, after mathematization has revealed formal equivalencies or isomorphisms between two competing approaches, certain philosophical problems connected to them become simply meaningless. For instance, the problems of quantum mechanics can be expressed either by the apparently deterministic Schrödinger equation or by Heisenberg's completely probabilistic and abstract calculus. Since von Neumann could prove that both formulations are isomorphic, the philosophical controversy about determinism can probably be settled in an unphilosophical way. This does not exclude a difference in heuristic content that might become poignant outside the domain of quantum mechanics where von Neumann's uniqueness theorem fails. Second, mathematization makes it possible to formulate some sophisticated 'logical cycles' within and to find the absolute limitations of a theory.

> [In the field of quantum mechanics,] by the best descriptions we can give today, there are absolute limitations to what is knowable. However, they can be expressed mathematically very precisely, by concepts which would be very puzzling when attempted to be expressed by any other means. Thus, both in relativity and in quantum mechanics the things which cannot be known always exist; but you have a considerable latitude in controlling which ones they are. … This is certainly a situation of a degree of sophistication which it would be completely hopeless to develop or to handle by other than mathematical methods. (Ibid., p. 487)

If string theorists are right to believe that there exists a final theory of physics, mathematics should rather make available a proof that there do not exist such limitations.

4. SOME LESSONS FOR STRING THEORY

The discovery of smaller and smaller subatomic particles of higher and higher energy and the general acceptance of big-bang cosmology has led to a rather peculiar picture of physical theory. The tiniest parts of the Universe are governed by the most fundamental laws, invigorated in the earliest split seconds of its existence. Although in virtue of symmetry breaking not all features of lower energy theories can be completely deduced from the

respective higher theory – that is, theory reduction fails at places –, most present-day particle physicists believe in a sequence of (ontological) reductions to ever more fundamental levels of reality. String theorists, additionally, hold that this series comes to an end at the Planck scale, such that strings represent the fundamental building blocks of nature. It seems evident that if string theory truly is the final theory of physics, it should not rest on mathematically shaky arguments. Or put differently, it cannot stay 'theoretical' forever.

But there is also a more specific problem that concerns the axiomatic expression of finality. Steven Weinberg, for instance, holds that the "final theory … is so rigid that it cannot be warped into some slightly different theory without introducing logical absurdities like infinite energies." (Weinberg, 1993, p. 12) Thus, it is logically isolated. "In a logically isolated theory every constant of nature could be calculated from first principles; a small change in the value of any constant would destroy the consistency of the theory." (Ibid., p. 189) Phrased in the language of Lakatosian quasi-ontology, there is no longer any basis to retransmit falsity to a single axiom; one would have to abandon them altogether in the face of striking counterevidence. The theory is thus immune against heuristic falsification. But this characterization of finality yields a problem. Why should physical heuristics, that is, ontological reductionism, and mathematical heuristics pull in the same direction. Or put differently, what is most fundamental to the theoretical physicist need not be most fundamental to the theoretical mathematician.

It is true, string theory exhibits a remarkable uniqueness. It can be consistently formulated only in 10 dimensions (for the fermionic string), there are no free parameters, and it automatically produces a smallest scale (Planck length). But in recent years, a large variety of dual string theories has emerged despite these strictures. Two dual theories have the same empirical content, but involve different basic objects and different topologies. Dawid (2003) identifies dualities as a source of problems for a realist interpretation of string theory. More generally, the duality problem is a somewhat paradoxical feature of the purported endpoint of a research programme motivated by ontological reductionism. Weinberg's finality criterion does not isolate only one single theory.

But even if there are no explicit ambiguities, it is always possible that a given theory, even the final one, may well be formalisable in two different axiom systems one of which is preferred by the physicist on ontological grounds while the mathematician cherishes the structural features of the other one. For instance, string dualities may turn out to be a deep mathematical fact, physicists' ontological quibbles notwithstanding.

To my mind, the problem is of a generic kind and reaches back to Hilbert's notion of 'deepening the foundations' which – ensuing from the analysis of the axioms' independence – was the heir of the ancient attempts to prove the fundamental presuppositions of science themselves. Within the axiomatic method it rather corresponds to an architectonic reorganisation of the axiom system. One can distinguish deepenings of different scope (Stöltzner, 2002c), the simplest one being just to drop a dependent axiom. Hilbert lauded Boltzmann and Hertz for having deepened the foundations of Lagrange's mechanics containing arbitrary forces and constraints to either forces without constraints or constraints without forces. Both deepenings expressed starkly different physical ontologies. Moreover, mathematical deepening occasionally arrives at a formulation hardly any physicist is familiar with, such as basing classical mechanics on Bertrand's maximum principle.

Hilbert's formulation of general relativity amounted to the strongest type of deepening because he attempted to reduce all physical constants to geometrical ones. In its aspiration to eliminate physical constants, it corresponded to string theory. Hilbert's work was clearly an instance of 'theoretical mathematics', while Einstein's was a model episode of theoretical physics. In a lecture at Göttingen, Einstein had sketched the open problem in his theory, and Hilbert worked arduously to solve it first. Following the line of thought of his earlier works in mathematical physics, Hilbert came out first with an action principle while Einstein presented a differential equation – which Hilbert, after getting to know his competitor's solution, tacitly inserted into the galleys of his paper. Hilbert's work was theoretical insofar as he found a new mathematical structure inspired by physics, the Hilbert action integral. It was also theoretical in the negative sense because the main theorem of the paper was not proven; even worse, it was flawed and thus the programme to reduce all physical constants to purely geometrical ones failed. The claim disappeared from later versions of Hilbert's paper where he cited a result which is today known as Noether's second theorem. But it is certainly not Hilbert's theoretical style that is responsible for the poor recognition of his approach among present-day theoretical physicists. On the basis of the concepts at his disposal, Hilbert spotted the deepest mathematical structure of relativity theory. Yet it did not agree with what physicists, rightly on their part, considered as the core structures of general relativity, to wit, the metric and the affine connection.

This shows in conclusion that even if one tries to make proper space for 'theoretical mathematics' by deliberately blurring the boundary between mathematical and physical ontology in favour of a Lakatosian quasi-ontology, one does not enter into a unique (platonist or realist) world. This might be bad news for some advocates of a 'Theory of Everything', but good

news for those who endorse a pragmatic view about the axiomatic method that emphasises its experimental character alongside its justificationary force.

REFERENCES

Aberdein, A., 2003, The informal logic of mathematical proof, arXiv:Math.HO/0306298.

Atiyah, M., et al., 1994, Responses to 'Theoretical Mathematics: Toward a Cultural Synthesis of Mathematics and Theoretical Physics', *Bulletin of the American Mathematical Society* **30**: 178-211.

Colyvan, M., 2001, *The Indispensability of Mathematics*, Oxford University Press, New York.

Corfield, D., 1997, Assaying Lakatos's philosophy of mathematics, *Studies in History and Philosophy of Science* **28**: 99-121.

Corfield, D., 2003, *Towards a Philosophy of Real Mathematics,* Cambridge University Press, Cmabridge.

Davey, K. J., 2003, Is mathematical rigor necessary in physics?, *The British Journal for the Philosophy of Science* **54**: 439-463.

Dawid, R., 2003, Scientific realism in the age of string theory, Preprint, Philsci-archive #1240.

Horgan, J., 1993, Culture clash. Is mathematics becoming too much like physics?, *Scientific American* August 1993, p. 14.

Jaffe, A., and Frank Q., 1993, 'Theoretical Mathematics': Toward a cultural synthesis of mathematics and theoretical physics, *Bulletin of the American Mathematical Society* **29**: 1-13.

Iagolnitzer, D., ed., 1995, *Proceedings of the XIth International Congress of Mathematical Physics*, International Press, Cambridge, MA.

Jaffe, A., and Quinn, F., 1994, Response to comments on 'Theoretical Mathematics', *Bulletin of the American Mathematical Society* **30**: 208-211.

Jaffe, A., 1997, Proof and the evolution of mathematics, *Synthese* **111**: 133-146.

Köhler, E., 2002, Gödels Platonismus, in Bernd Buldt et al., eds., *Wahrheit und Beweisbarkeit*, vol. 2, Vienna: ÖBV-HPT, pp. 341-386.

Lakatos, I., 1976, *Proofs and Refutations. The Logic of Mathematical Discovery,* edited by John Worrall and Elie Zahar, Cambridge University Press, Cambridge.

Lakatos, I., 1978a, *The methodology of scientific research programmes*, Philosophical Papers Volume 1, John Worrall and Gregory Currie, eds., Cambridge University Press, Cambridge.

Lakatos, I., 1978b, *Mathematics, science and epistemology* (Philosophical Papers Volume 2), John Worrall and Gregory Currie, eds., Cambridge University Press, Cambridge.

Leng, M., 2001, What's wrong with indispensability?, *Synthese* **131**: 395-417.

Maddy, P., 1997, *Naturalism in Mathematics,* Oxford University Press, Oxford.

Stöltzner, M., 2001, Opportunistic axiomatics – John von Neumann on the methology of mathematical physics," in M. Stöltzner, M. Rédei, eds., *John von Neumann and the Foundations of Quantum Physics*, Dordrecht: Kluwer, pp. 35-62.

Stöltzner, M., 2002a, What Lakatos could teach the mathematical physicist", in: G.Kampis, L.Kvasz, M. Stöltzner, eds., *Appraising Lakatos – Mathematics, Methodology, and the Man,* Dordrecht: Kluwer, pp. 157-187.

Stöltzner, M., 2002b, Bohm, Bell and von Neumann. Some philosophical inequalities concerning No-go Theorems, in T. Placek, J. Butterfield, eds., *Modality, Probability, and Bell's Theorem,* Dordrecht: Kluwer, pp. 35-56.

Stöltzner, M., 2002c, How metaphysical is 'Deepening the Foundations'? Hahn and Frank on Hilbert's axiomatic method", in: M. Heidelberger, F. Stadler, eds., *History of Philosophy of Science. New Trends and Perspectives*, Dordrecht: Kluwer, pp. 245-262.

Stöltzner, M., 2004, On optimism and ppportunism in applied mathematics (Mark Wilson meets John von Neumann on mathematical ontology), *Erkenntnis* **60**: 121-145.

Thurston, W., 1994, 'Theoretical Mathematics': Toward a cultural synthesis of mathematics and theoretical physics, *Bulletin of the American Mathematical Society* **30**: 161-177.

Tisza, L., 1997, The reasonable effectiveness of mathematics in the natural science, in *Experimental Metaphysics,* Robert S. Cohen et al., eds., Dordrecht: Kluwer, pp. 213-238.

von Neumann, J., 1932, *Mathematische Grundlagen der Quantenmechanik,* Berlin: Springer; English translation by Robert T. Beyer *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton, 1955.

von Neumann J., 1947, The Mathematician, *Collected Works,* vol.1, pp.1-9; originally in R.B. Heywood, ed., *The Works of the Mind,* University of Chicago Press, Chicago, 1947, pp. 180-196.

von Neumann, J., 1954, The role of mathematics in the sciences and in society", *Collected Works*, vol. 6, pp. 477-490; originally in Princeton Graduate Alumni, June 1954, 16-29.

von Neumann, J., 1955, Method in the physical sciences, *Collected Works,* vol.6, pp. 491-498; originally in L. Leary (ed.): *The Unity of Knowledge*, New York: Doubleday, 157-164.

von Neumann, J., 1961-1963, *Collected Works*, 6 vols. Ed. by Abraham H. Taub. Pergamon Press, Oxford.

Weinberg, S., 1993, *Dreams of a Final Theory*. Vintage, London.

Wigner, E. P., 1960, The unreasonable effectiveness of mathematics in the natural sciences, *Communications in Pure and Applied Mathematics* **13**: 1-14.

Wilson, M., 2000, The unreasonable uncooperativeness of mathematics in the natural sciences, *The Monist* **83**: 296-314.

PART 3

FINAL REMARKS

MATHEMATICS AND PHYSICS

VLADIMIR I. ARNOLD
Steklov Mathematical Institute, Moscow, Russia

Abstract: In this paper I present, through examples, the problem of the mathematical rigour of the bases of physics and explain what the utility of a precise mathematical perspective of the real world is. I also offer some arguments for the existing difference in the approach to the truth as understood by mathematicians and physicists.

Key words: mathematical rigour, physics; experimental science.

The word "mathematics" means "precise knowledge". Barbaric peoples, with no inclination for such things, had no corresponding words in their languages, so that now in almost all languages one uses the uncomprehended Greek term. The only exception to this is the Dutch language, for which Stevin already in the seventeenths century fought against the pollution of the terminology by foreign words, and insisted in the translation of all terms into mother tongue words. So, the term "viskunde" - i.e., "knowledge" -, since childhood brings mathematics close to the real world.

When Ya. B. Zel'dovich, eminent theoretical physicist and one of the founders of Russian nuclear physics, gave birth to his *Higher Mathematics for Beginners Physicists*, he raised the terrible anger of the Russian mathematical literature censor of the time - the Academic of Sciences L.C. Pontryagin.

He rightly showed that in his book, Zel'dovich had defined the derivative of a function *as* "*the quantity* expressing the ratio between the function increment and the argument increment, provided that the last one is small".

The mathematician was indignant at the complete exclusion, in this definition, of concepts of the limits theory, as well as of a considerable part of the logical bases of the mathematical analysis, which attained his

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 225–233.

perfection only at the end of the nineteenth century, with the construction of a coherent theory of the real numbers continuum.

Zel'dovich answered in this way: We are always interested only in ratios *of finite* increments, and never in any abstract mathematical limit.

To take the argument increment - say, of the coordinate of a point or of the time moment – less than, say, 10^{-10} or 10^{-30} (in reasonable measurement units) – “evidently exceeds the model precision because the *physical structure of space (or of time) inside such small intervals already do not correspond to the mathematical theory of the real numbers* (*as consequence of the quantum phenomena).*”

“The question consists simply in the fact - Zel'dovich continued - that to find the finite increments ratios which interest us is difficult; for this reason approximating *asymptotic formulae* was invented for them. Mathematicians call these approximating asymptotic formulae by their words ‘limits’ and ‘derivatives’. In any real application of the theory one must consider increments sufficiently small as to have a correspondence of the theory with experiments, but smaller increments are not needed.”

The long discussion had as consequence that Pontryagin wrote his own textbook on the analysis principles. Already in the introduction of this book, Pontryagin indicated that “some physicists believe that it is possible to study and to apply analysis, avoiding its absolutely logical notations, and the author of the present textbook … agrees with them”.

I was remembered of this discussion on the mathematical rigor of the bases of Science, when my close friend M. L. Lidov, who was working on the calculation of trajectories of Sputniks and of the space-shuttles, began to quarrel with me about my course on the theory of differential equations (at that time he gave a course at the Moscow State University on the ballistics of Sputniks, and we often discussed together, mainly because at that time I was myself working in celestial mechanics).

“As all mathematicians - told me Misha - you teach the uniqueness theorem, according to which *the integral curves of the ordinary differential equations do not intersect.* But this statement (albeit you prove it correctly and perfectly) is untrue. For example, the equation $dx/dt = -x$ has solutions $x = 0$ and $x = e^{-t}$. The integral curves are the graphs of these functions (any computer can draw them) and you see that they are clearly intersecting each other. Indeed, for example, at $t = 10$ between any two of such integral curves even one atom cannot pass.

So, the uniqueness theorem is only a mathematical fiction, having little to deal with the real world.”

After that, the interlocutor explained to me that it is exactly for the effect above that, on landing, at the last moment the seaman throws the rope on the quay, where is quickly secured to the bollard (sometimes, this is done by the

seaman himself, who jumped on the quay). Finally, the last part of booring is hand made by winding the rope.

This has the following explanation. Automatic landing, corresponding to the general principles of control theory, is based on the *negative feedback.* Depending on the distance x remaining to the landing, the control is done in such a way that the velocity decreases to zero (as a function of x). Of course, this function is smooth, i.e., for small distances the velocity will vanish with x approximately linearly.

According to the theorem of uniqueness mentioned above, *the booring time will be infinite for any smooth feedback mechanism.* To land in *a finite time,* one has either to renounce to the regularity principle (with a smooth negative feedback), changing the control of the boat velocity by the work the seaman does on the rope, or accept that the boat strikes the quay (it is for this that worn out car tires hang on the dock).

The fact that all this is not discussed by mathematicians neither in courses of theory of dynamical systems nor of differential equations, nor in theory of control and optimisation, is, of course, a displeasing consequence of the long lasting detachment from the real world, from physics and technics, of mathematicians who live in the ivory tower of their axiomatic science.

M.L. Lidov knew very well the axiomatic science, but he was interested in the above problems because he was dealing with the calculation of the landing of the space shuttles on the moon, where one encounters the same problems as in the landing of boats.

Since I don't want to bound myself only to criticism, I give another example of the great utility of the precise mathematical view point of the real word, taking it from another work by Lidov.

The Moon turns around the Earth along an orbit lying nearly in the ecliptic (i.e., in the plane of the orbit of the Earth around the Sun). The famous "*Laplace Theorem on the stability of the Solar System*" states that if the inclination of the Moon orbit on the ecliptic is small, then, neglecting the perturbation due to the Sun influence, the Moon orbit will slightly oscillate (giving rise to eclipses), but it will not change systematically (neither falling to the Earth nor going away).

Lidov posed to himself the problem of what should happen if the initial Moon orbit were *strongly inclined on* the ecliptic - say - making an angle of 80 degrees with it (while staying at today's distance from the Sun).

Of course, it is impossible to force our Moon to move in such an orbit. However, it is possible to put an artificial satellite on an orbit perpendicular to the ecliptic. The question on the evolution of its orbits (under the Sun attraction) is of real interest for the Sputnik future.

Lidov’s result was very surprising: *Such artificial moon would fall on the Earth in three years*. Therefore, it is not convenient to put a satellite on such an orbit.

The reason of the falling does not consists in the vanishing of the orbit radius (the mean distance of the satellite from the Earth), but in the reduction of the shortest axis of the ellipse along which the satellite is moving - i.e., in the increase of ellipse's eccentricity.

Even if the initial orbit of the artificial moon where with good approximation like a circle, the perturbation would quickly transform it into an ellipse (with the shorter axis decreasing in time). Whereas the highest axis of this ellipse should keep its length (as Laplace proved) equal to the diameter of the non-perturbed orbit (i.e., the diameter of the today's Moon orbit) the increase of eccentricity in time would make this narrow ellipse finally similar to a segment (travelled forward and backwards).

As a consequence of the big eccentricity, the orbit of the artificial moon would begin to intersect the Earth, so that such satellite should fall on the Earth, whereas its *mean* distance from the Earth centre over one revolution period should remain equal to the same mean distance of today's Moon (even at the very moment of falling).

A few words, now, on the difference of opinions between physicists and mathematicians on the character of our common science. At the end of the Second Millennium of our era, the journal *Uspekhi Fizicheskikh nauk* (Russian Physics Surveys) published a jubilee issue and asked me to write for this issue a survey “Mathematics and Physics” (two other mathematical articles on the same journal where written by K. Weierstrass and C. Jacobi).

What struck me, was that the journal editor erased from my article two clear demonstrations of the strong difference between the approaches to the truth understanding by physicists and by mathematicians: one of these demonstrations was contained in a citation, chosen as epigraph, from the book by E. Schrödinger on thermodynamics, and the second was contained in a problem for children.

These are the passages, evidently non understood by the editor. There were in my article two epigraphs. The first one (kept) was a statement by Stendhal: “*Among all sciences I like mathematics the most, because in this science any hypocrisy, which I most detest, is totally impossible*”. It seems that Stendhal liked the fact that in mathematics, if it was somehow calculated that seven by six is equal to forty two, this would hold forever: the truth is final and unquestionable.

Schrödinger instead wrote: “Let us suppose that α is equal to zero, even though, firstly, α cannot be equal to zero, and, secondly, vanishing α contradicts the quantum mechanics”. Evidently, physicists prefer not to

make apparent their constant hypocrisy, with their ambiguous terminology and the internal logical contradictions of their theories.

Afterward, when I tried to discuss the evident differences with the academic V.L. Ginzburg, the journal editorial chief, he demonstrated to me that "*mathematicians in general cannot understand anything in physics*", showing me a formula in his own article. "Which in your mind is the meaning of these symbols?" - he asked.

Thinking to have understood, I answered: "Index *i* is repeated, so, according to the Einstein convention, it is a positive definite form - a sum of squares. Only I don't know how many they are, because the sum limits are not specified".

Well, - cheered the physicist - as all mathematicians, you do not understand anything. Letter *i*, you see, is 'Latin', and not 'Greek'. This *means* that its values are four: 0, 1, 2, and 3. As for the *sum*, this is absolutely not the case: this notation is relativistic; therefore one of the squares must be taken *with opposite sign with respect to the others.*

I did not succeed in persuading my interlocutor that it is not appropriate to indicate a subtraction with the symbol of sum (and that a limit like "velocity not higher than 60" is a nonsense while it is not specified whether one means kilometres per hours or parsecs per second).

But there is now the second example, showing the cardinal difference of way of posing and understand problems by mathematicians and physicists.

In my article there where two examples (taken from old text-books). Mathematical question: "*On a book-shelf there are two volumes of Pushkin 's poetry. The thickness of the pages of each volume is 2 cm and that of each cover 2 mm. A worm holes through from the first page of the first volume to the last page of the second, along the normal director to the pages. What distance did it cover?*"

I gave the unexpected answer: 4 mm. The journal editors thus corrected it into: "from the *last* page of first volume to the *first* page of the second". The topological thinking is more difficult of what one may expect from the editors of a physics journal. Journal editors try always to change into the usual triviality any statement having originally an opposed meaning.

Another example of the typical physical style is given by the following problem taken from an old textbook.

A man of S. Petersburg went rowing along the Neva River against the stream. When he was under the Troitski Bridge, he lost his hat. Reaching the Liteinyi Bridge, he met a friend, who informed him of this loss. Then the man went back with the same speed with respect to the river stream as before and got his hat after 20 minutes, under the Dvortsovyi Bridge. Find the velocity of the Neva stream.

For mathematicians it is evident that this problem is not solvable. However, with the proper physicists hypocrisy the solution was provided: following the Galileo principle of relativity, the man rowed off from the hat against the stream and reached it boating in the stream direction in equal time intervals of 20 minutes. This means that the hat went from the Troitski to the Dvortsovyi Bridge in 40 minutes. Since the distance between these bridges is one mile, then...

All physics textbooks are written in this style: there are not plainly expressed some distances between bridges or other things speaking of which "does not matter".

The mathematical rigor is often attained with difficulty even from good mathematicians. The following example is taken from the famous book by Courant and Robbins *What is Mathematics*.

> Let us suppose that a wagon is moving along a horizontal track, and that a rod, with one end hinged to the wagon's floor, may rotate around a fixed horizontal axis orthogonal to the railway.

The statement is that for whatever given motion law of the wagon (in the time interval from zero to one) the initial position of the pendulum can be chosen in such a way that in the final instant it will not be horizontal. (This problem was suggested by H. Whitney).

The authors demonstrate this in this way: if the initial position of the pendulum is horizontal and in the motion direction, then it remains horizontal. If the initial position is horizontal in the opposite direction, then it remains horizontal. Consider now an arbitrary initial condition. The initial one defines the final position. This is a continuous function taking the values "forwards" and "backwards". By a topology theorem, it takes also all the intermediate values, which completes the proof.

Some years ago I was requested from professor Robbins (Courant was already dead at that time) to try to improve this "incorrect proof". In fact, it is not immediate to see any continuous function 'final position versus initial position'. One must define it exactly (taking into account the influence of the allowed hits to the wagon) and demonstrate its continuity.

I heard that some American mathematicians, trying to do this, wrote a demonstration with erroneous intermediate statements so that the problem of the rod even today appears to be open[85].

In a meeting of the French Academy of Sciences, I told that "*mathematics is a part of physics, being, as physics, an experimental*

[85] Discussions on this problem are published in (Blank, 2001; Gillman, 1998; and Littlewood, 1986).

science: the only difference is that experiments in physics cost usually millions of dollars, whereas in mathematics they cost a few cents".

An eminent French mathematicians sent me a letter, in which he wrote that, on the contrary, "mathematics has nothing to share with physics".

Some time later in an official debate on the education problem in Moscow, academician D.V. Anosov intervened with the following "criticism of Arnold": Arnold put (and this is true) in his paper "Polymathematics: is mathematics a single science or a set of arts?" in the book *Mathematics: Frontiers and Perspectives* (Arnold, Atiyah; Lax; Mazur Eds.) the comparison of opinions of two great algebrists. Hilbert, in 1930, in the article "Mathematics and Natural Sciences" writes that "*Geometry is a part of Physics*", while the above-mentioned french mathematician claims that "*Mathematics and physics have nothing in common*".

In these two statements Arnold - the lecturer said - sees a contradiction. The reason of this is that Arnold, due to his intellectual lacks, either did not read, or did not understand Aristotle. Indeed, in my mind, having read and understood Aristotle, there is no contradiction in this, because there is the consequence: *mathematics has nothing in common with geometry.* For this reason - in this way the academician terminated his speech - I propose to *eliminate completely geometry from all mathematical courses* (in universities, in high school, in junior high school, in elementary school).

Few weeks later I received from the Russian Minister of Education the Ministry project of new programs for schools in all subjects. Following the Anosov's opinion, *geometry courses were completely eliminated from all education programs.*

Afterwards, I was fighting against this obscurantist decision; letters against the elimination of geometry were sent to the Ministry from the Scientific Council of the Steklov Mathematical Institute of the Russian Academy of Sciences and, on the other hand, from the representatives of several war industries (informing me about this a year later in Dubna). After some months, the Minister sent me (with his thanks) the new elaborated version of the education programs, where geometry had been returned to his old place[86].

Another difference between physicists an mathematicians was remarked by Nikolai N. Bogolyubov, the director of the Mathematical department of the Russian Academy of Science, who always tried to persuade me to

[86] A manager of a supercomputers factory recently wrote: "Geometry has to be transferred into history courses, because all problems of it are either solved or can be solved with other methods" (Bailey, 1996). Any attempt to explain to such people thinking, logic, esteem of science and culture is hopeless.

publish my articles not in mathematical, but in physical journals. According to his words, the number of readers of a good paper will be the same, say, one thousand. “The difference - he continued - consists in the fact, that after a publication in a Journal of mathematics, these thousand readers live in a century, i.e., ten readers per year, and this is the eternal glory. After publication in a journal of physics, all these thousand persons read the paper in few weeks, and the author is immediately elected member of Academies, but one hundred days later nobody will remember the name of the author, whereas the results and the methods contained in his article will be continuously used by all, as common knowledge (and, surely, without citation of the author and with a consequent award the Nobel prize for his discover to other people)”.

I remember also that N.N. Bogolyubov showed me a wonderful example of the advantage of his pragmatic point of view. At that time I wanted to publish the Russian translation of the Poincaré selected works, but the editor refused (citing the critic of Poincaré, published in 1909, in *Materialism and Empirocriticism*). When I asked N.N. Bogolyubov, who had developed Poincaré’s ideas, to help me, he said: “We shall use the fact that Poincaré, as both you and me, was not only a mathematician, but also a physicist, even a naturalist. But a naturalist must see in any natural phenomenon, even unpleasant, as the volcanoes eruptions, the possibility to utilize it for scientific purposes, for example, to know something on the internal of Earth.

In our case we are dealing with another unpleasant phenomenon of nature, that we need to utilize: it is the antisemitism and the anti-eisteinism of various people”.

Saying this, he wrote a letter to the editor, explaining (in all fairness) which big merits Poincaré had in the foundation of Relativity. He published the relativity principles in his article “On the measure of time” ten years before Einstein, who only in the forties admitted, under the advice of his teacher, Minkowski, to have examined the Poincaré works from the beginning of his own.

So, three volumes of the Poincaré selected works were published in Russian, including the article on the measure of time, but without any Einstein criticism.

REFERENCES

Bailey, J., 1996, *After Thought*, Basic Books.

Blank, B., 2001, Book review: What is mathematics?, *Notices of the Amer. Math. Soc.* **48 (11)**: 1325-1329.

Gillman, I., 1998, Book review: What is mathematics?, *Amer. Math. Monthly* **105 (5)**: 485-488.

Littlewood, J. E., 1986, *Littlewood's Miscellany*, Cambridge University Press, Cambridge, pp. 32-35.

VALUES AND MEANING IN THE QUANTUM UNIVERSE

NIKOLA ZOVKO
Ruđer Bošković Institute, Zagreb, Croatia

Abstract: Concepts of a natural set of values and a natural (intrinsic) meaning within a quantum Universe are discussed. Going beyond the Copenhagen interpretation orthodoxy by the Dirac-Heisenberg-Pauli-Stapp ontological model, it seems possible, at least in principle, to have a philosophically acceptable material Universe that is also inherently meaningful.

Key words: natural ethics; Universe; quantum mechanics.

1. INTRODUCTION

After stating that the last century was the bloodiest and the most cruel century in the history of mankind, the recent UNESCO Conference on Science in Budapest declared the 21st century to be the century of ethics. But we know what is going on just at the beginning of this century. The situation seems desperate and frustrating.

Ethical fundaments of global religions, although rather similar, are of no help: they more divide and cause wars then they unify and create peace. Ecumenism is practically dead.

Is the scientific approach of much practical help? More than ever it is needed today. To which extent can the science offer an empirical basis for a philosophically acceptable selfcontained matterial universe and speak of a meaninful universe at least in principle? “More we understand it more it seems pointless” is the famous Weinberg's discouraging claim. Can one take a more optimistic position having in sight a meaningful quantum universe?

G. Boniolo et al. (eds.), The Role of Mathematics in Physical Sciences, 235–244.

Since concepts of absolute set of values and absolute meaning are illusions, how about asking the Nature herself about some natural values and natural meaning? One might hope to read them off from the world view emerging from quantum properties of the Universe. Following Henry Stapp, the answer to those questions is positive. At a fundamental level the Nature indeed makes “decisions”, generates “choices” and stands “consequences”.

Generally, question of ethics is glued to the Cartesian problem of the mind – matter orthogonality, more than three centuries. It is well understood today that the realm of classical physics has no natural place for human mind, not even for the very life. No any set of values can be attributed to the classical universe and humans are scientifically justified automata obeying mathematically expressed cold laws of deterministic classical physics. Being predetermined, they are responssible for nothing and their ethics consists merely of their own interests and survival. Of course, today we are aware of the fact that even the classical physics is not really fully deterministic due to essential uncertainties in the classical knowledge, but the prevailing natural phylosophy is considered to be deterministic. The only ethical act in this classical world view is the choice of initial conditions that fully fixed the destiny of the entire spacetime. This is reserved for God.

At the other hand, quantum theory creates, at least in principle, the possibility of a fundamental bridge between the matterlike and the idealike things in nature[87]. A great deal of physicists now tries to look beyond the ortodoxy of the standard interpretation of quantum theory, which is pure epistemology, and eventually asks “What is really happening there?”. As will be disscussed later, at the level of actual quantum events (in what follows called Heisenberg events) one may recognize that a profound quantum choice takes place everywhere and forever and injects the meaning into the physical universe. Menthal universe is subject to the same mechanism once we accept the idea that human conscious thoughts are just the actual quantum events over the entire brain or over a large part of it. Thus the menthal and the matterial universes are brought together on a deeper level of physical reality beyond our direct experience. Such a realm can in principle accommodate ethical concepts of choice, meaning, value, etc.

[87] An excellent introduction into the subject may be found in (Stapp, 1993)

2. CLASSICAL MECHANICS AND ITS STILL PREVAILING WORLD-VIEW (WELTANSCHAUUNG)

As mentioned in Introduction, classical mechanics fails badly when we deal with the mind-matter problem, the problem of ethics and the like. Why is it so, and what is the basis of the natural philosophy fixed by the classical mechanics?

All motions in the Universe are fixed by deterministic differential equations generated by the mathematically expressed lows of nature. In order to solve those equations completely one has to know the state of motions (initial conditions) at some earlier time, usually shifted to the "beginning of time", to the time the Universe was created. But who did choose this initial condition? If, by definition, a choice means a fixing of any aspect of nature not fixed by known laws of nature, then the classical physics contains only one such choice – "ethical act". This also holds for motion and behaviour of humans in time. They are fully predetermined; they cannot do anything by their free will; whichever way they behave science gives them a full excuse for their eventually unethical behaviour. Moreover, classical physics cannot define anything that would look like a set of values in accord with which humans should behave. Even if such a set of values could be given, then the role of science is understood as serving to reach these values, not to define them. So, if a man/woman would behave in accord with some prescribed ethics for benefits of the entire society, his/her behaviour would still be explained as doing well for the society because he/she is convinced it is the best for his/her private interests. His/her behaviour is entirely determined and led by own interests since there is no scientific foundation of what is a value.

Finally, and most important of all, is that classical mechanics failed badly in describing the material universe in the micro world. All atomic and subatomic phenomena did not follow its predictions. The fresh new ideas about the concept of the physical reality had to be accepted. Certainly, physical reality is not that what we observe by the direct perceptions of our senses. It just happened so that the phenomenon of life and the human mind appeared at a certain (classical) scale in the material universe at which its quantum properties do not get so drastically pronounced. In short, we can say that the world-view based on classical physics provides a scientific justification for eventually unethical behaviour of human beings. As also in few other disastrous cases in history of physics, the situation was saved by quantum mechanics.

3. QUANTUM MECHANICS AND THE WORLD-VIEW EMERGING FROM IT

What did the quantum theory? Besides of giving us a consistent picture of our understanding of the material micro world, it opened the window for the explanation of the existence of life (Wigner, 1967; Bohr, 1934, 1963), it offered us a bridge to connect matter and mind (Stapp, 1993; Wigner, 1962), and it gave us a real chance to shape some natural ethical concepts consistent with the quantum vision of man and its role in the quantum universe (Stapp, 1993). The essential feature of quantum laws is that they are of statistical and nondeterministic nature. According to orthodox thinking, they fix not what in nature actually happens, but only probabilities for various things that might happen. Quantum theory is a two-component structure sketched by Roger Penrose as (Penrose, 1994)

$$\text{Quantum Theory} = U + R.$$

Here U represents a unitary deterministic, local, linear and time symmetric evolution of the quantum system as given by some quantum evolution equation (like the Schrödinger's equation), while R represents what is called reduction[88] or collapse of the wave function – source of all troubles with the interpretation of quantum mechanics. In all aspects R is opposite to U: it is nondeterministic, it is nonlocal, it is non-linear, and it is nonsymmetric in time. In connection with it one talks also about very unpopular quantum jumps of a system. Since the beginning of quantum theory they sit on nerves of many physicists. Especially hard to accept is the property of nonlocality (Bell, 1987). Since those early days up to the present times a huge activity, both theoretical and experimental, is concentrated on that problem.

The borderline between U and R is governed by some unknown physics one can only speculate about. According to Werner Heisenberg (Heisenberg, 1958), each quantum jump is a "choice" or a "decision" that picks out and actualises just one out of many linearly superposed possibilities previously generated by the unitary evolution process U. It is the process R that brings into the theory the unnatural element of chance and looks as a hand of God that, within the complex structure of possible physical realities generated by the process U, selects the reality actually appearing to us. Since this is of fundamental importance, we will elaborate this in more details in what follows.

[88] The most recent review of the problem of reduction of the wave function with an exhaustive list of references may be found in (Bassi and Ghirardi, 2003).

First, immediately after any observation, a quantum system does not exist anymore as that what was observed but rather evolves in time as a complex linear superposition of different quantum states. Each of those alternatives in linear superposition is fully determined by the evolution equation and being unitary means that the sum of all their probabilities at any time equals to unity. How long lasts this soup of alternative possibilities? Until the next observation, or better, until the next registration by a measuring device (photographic plate, Geiger counter and the like) is done. In that moment the quantum jump or a quantum choice takes place.

All paradoxity of a quantum jump is illustrated with the following example. Assume that a spin zero particle is emitted from a centrally symmetric quantum source. The space is isotropic and homogenous. Since there is no preferred direction in space, the wave function depends only of the distance and has a constant value all over an arbitrarily chosen sphere. Geiger counters are spread over the sphere and we know only one of them will make a click and register the emitted particle. When the counter makes a click, the quantum system jumps and the wave function suddenly reduces (collapses) over the sphere and assumes a form essentially different from zero only around the position of the counter that has registered. A sort of information travels in zero time (nonlocality) over the sphere of whatever size we can imagine. Strange indeed behaves the Nature in the micro world! Or should this refer only to our knowledge?

Why should we call it a quantum choice? Because the Nature makes a choice when the system jumps into the state just being observed. Henry Stapp calls it a Heisenberg event (Stapp, 1993) and it represents a fundamental element of the physical reality. After the event again a soup of superposed states evolves until the next Heisenberg event takes place and so on. One may visualize a picture of two intertwined chains (U and R).

Terms “quantum choice” and “Nature makes a quantum choice on her part” were first used by Paul Dirac. And so, by continuously making quantum choices always and everywhere, Nature injects into the Universe the entire particularness that we observe. Nature is doing an “ethical act” over the entire space-time manifold. Of course, all this holds within what we call the Dirac-Heisenberg-Pauli ontologicalization of quantum mechanics.

Asking the central questions “What actually happens there?” and “Who governs the quantum choice?” leads us to several different proposals. The most advocated one is the Copenhagen interpretation, which is pure epistemology. It does not describe quantum systems; it describes our knowledge about quantum systems. Albert Einstein called it a “soft pillow” for physicists, while David Mermin named it a “shut up and calculate” (transition probabilities) proposal. It is a closed door for any deeper philosophical insight. Many consider it as a brain washing for generations of

physicists. Details may be found in recent textbooks discussing the interpretations of quantum mechanics. Also popular is parallel minds (universes) model, and pilot wave model, but seemingly they create more problems than they solve. In what follows we stick to the Dirac-Heisenberg-Pauli-Stapp proposal, just mentioned above.

In his famous book (Heisenberg, 1958) Heisenberg writes:

> …observation makes a change in the probability function abruptly; out of all possible events it selects the actual one that just has taken place…
>
> …transition from "possible" to "actual" happens during the act of observation…

It is therefore the observation (registration) that transforms "potentiality" into "actuality". A quantum system lasts in time in a variety of virtual states but only one out of those states becomes actualised by the very fact of observation. The idea of the potential and the real is rather old in natural philosophy (Boscovich, 1763).

4. UNIFICATION OF MENTAL AND MATERIAL UNIVERSES

It seemed for three centuries absolutely impossible to find, even in principle, a way to unify the matter like and the idea like objects in the world. Today this task seems equally impossible if we stay in the realm of classical physics and its world-view. But the so called Grand Unification of three fundamental forces in physics teaches us the following: if you cannot unify them at ordinary (classical) conditions where things look so desperately different, go to energies far away from those in our laboratories, and the things might appear similar in some respect. Indeed, the electromagnetic force has an infinite range, while the weak and the strong nuclear forces reach far shorter than a billionth part of a metre. Their strengths are also very different. How can one bring them together? As it is well known today, a way out is to go to energies of hundreds, millions or billions of GeV, very far from energies achievable in our laboratories. There those forces become similar and have the strengths of the same order of magnitude. Why should not one try to look for similarities of the mind and the matter in regions far away from the realm of classical physics?

Quantum mechanics, according to standard Copenhagen interpretation, has no answer to the mind-matter problem: the task of physics is to predict and not to understand, they claim. Among physicists belonging to the Copenhagen circle only Heisenberg was eventually willing to ask and put

questions like "What really happens there?". Henry Stapp, therefore, calls his unification model the Heisenberg – James model. He combines the Heisenberg's ontologicalization (Heisenberg, 1958) of quantum mechanics with observations in psychology done by William James more than a century ago.

William James discovered (James, 1890/1950) in psychology the essential feature of quantum mechanics. He found, namely, that the conscious thought has properties specific for what we today call a quantum object in atomic and subatomic phenomena! Either "whole or nothing" property is characteristic for a thought. One cannot cut a thought into two peaces that are also thoughts. The meaning is lost if we decompose a thought into its components, like the word looses its meaning if analysed in terms of its letters. This is the very property of a quantum system. Either the whole photon enters the counter or it will not be registered at all. The idea of William James was not accepted at that time because physicists did not yet arrive to the quantum mechanics.

Biophysics and neurosciences tend today to a general conclusion: if quantum transitions take place over the entire brain or a large part of it, then the human brain, in an important aspect, behaves as a quantum measuring device. John von Neumann was first who showed a half a century ago that quantum events in human brain need not to happen at the level of individual neuron firings or individual synaptic discharges (von Neumann, 1955). Quantum events are taking place over the whole brain in a correlation with occurrences of conscious thoughts. Conscious thoughts are quantum events in the brain!

5. ARROW OF TIME AND INTRINSIC MEANING

Deeper insight into ethical concepts brings us immediately to questions about the category of physical time: what is the true nature of time, why the arrow of time is fixed and the like. The concept of meaning is based on a definite arrow of time, on a sense of duration, on a direction with endurance. Meaning persists in time as a process that sustains itself and refines itself. According to Henry Stapp (Stapp, 1993),

Meaning = a mechanism that enables a form
to be recreated in a refined form.

Of course this mechanism is not dynamical one, it is an element of chance. Hence, endurance and reproducibility, essential features of meaningful forms, are nondynamically generated intrinsic properties. Such an intrinsic meaning is carried by a form if it persists in unidirectional time; it reproduces and refines itself without influence of some external agent.

Are there obvious carriers of meaning in the quantum universe? Yes, there are such carriers and they must be pure quantum states. They are characterized by local observable properties. Superposed states cannot carry any intrinsic meaning because the interaction with the environment destroys the property of endurance and reproducibility. Such states decay. Natural and exclusive carriers of meaning in the quantum universe are local observable properties associated with pure quantum states. The quantum law of evolution, after any observation, continuously creates a variety of forms that can act as carriers of meaning. Among them, through Heisenberg events, Nature chooses those that have the property to sustain and refine themselves. What can we read off from such a choice? Is there some obvious meaning in that what the Nature has chosen? We can take it only as the definition of the natural meaning that is very relevant for us if we are to survive in such a quantum universe. A full set of quantum states develops (Schrödinger's equation) after any observational act; any of them could become reality, but according to ideas of Dirac and Heisenberg, states actually chosen are those of an exceptionally special kind – those states are forms that last in time and sustain themselves.

How the Nature chooses events that will be actualised and become elements of physical reality? Things happen so as if She (mother Nature) considers each form having in view not what it is but what it does, how it behaves and what it produces in the quantum universe. Choice is "by purpose". Take as an example a number of protons closed in a box. As time is passing by they will assume a vast variety of different forms, and intrinsically all of them will be equivalent. Having in view only what they are, with no external agent, there will be no discrimination among them. However, there is one logical distinction among them, and this distinction is a very special one because it does not refer to any structure outside the form itself: one deals with the property of the form to sustain itself. Outside of known lows of physics a fundamental ethical act is continually happening in the quantum universe. In words of Henry Stapp (Stapp, 1993),

> … quantum choices are meaningful choices, where meaningful is defined intrinsically, within the quantum system itself, with no reference to any external criterion of meaning: it is defined in terms of the sustainability of the form…
>
> …quantum formalism is such that the quantum choice is a grasping, as a unified whole, of certain combination of possibilities that hang together as a local enduring form. Actualisation of this form utilizes and restructures some of quantum potentialities and produces an immediate rearrangement of possibilities for the next Heisenber event to occur.

Sustainable forms last only if the arrow of time is stable. Therefore, a meaningful universe must have a fixed arrow of time. However, all fundamental equations of physics are formulated with einsteinian time and are symmetric in time: they do not change if we replace t with –t. Equations stay the same if time starts running backwards. In order to induce the meaning into universe, God had to give a definite direction to time. How this was and still is one of greatest problems of fundamental physics. Where is that hidden place in our theoretical understanding of the history of Universe from which on the time starts to be undirected and the quantum nature of matter generates an intrinsically defined notion of meaning? So, a satisfactory foundation of natural ethics will seemingly wait for a deeper understanding of the nature of time (Penrose, 1994).

The very happening of quantum choices is spread over the entire space-time, and, therefore, it creates meaning locally where Heisenberg's events occur. However, the effect of the mathematical formalism of quantum theory is such that each event is registered globally. Quantum theory does not allow a free physical system to exist (Bell, 1987). Everything in the Universe is correlated. We elaborate this on the case of a single proton: When a Heisenberg event occurs and a detector registers a proton, this means that the potentiality for its detection is actualised in that detector while the potentialities for its detection in distant regions immediately vanish. Quantum world acts as a whole. The wave function collapses to zero everywhere except in the domain of the detector that made a click. Therefore, the quantum choice that occurred should be considered to be a local affair, because it actualises (brings into physical reality) a particular meaningful form in a local region of space-time. However, the collapse of the wave function took place globally, and the event is registered in a global bookkeeping: when the Heisenberg event occurred the rearrangement and adjustment of potentialities was immediately made over the entire Universe, mathematically over the entire space-time manifold.

6. CONCLUSION

Concluding this short review one may say that the conception of the quantum Universe emerges today naturally from the theoretical foundation and experimental verification of quantum mechanics. Among several models that try to explain what is the quantum physical reality, the most popular one seems to be the Dirac-Heisenberg-Stapp proposal. Their quantum ontology provides the answer to the question about fixing things not fixed by known laws of physics. As a consequence of this ontology one may say that (Stapp, 1993)

...under particular kind of conditions, Nature makes a choice ..... and locally induces meaning into Universe…

…the condition under which Nature acts is construed as an expression of a criterion of natural value.

Therefore, there is only one step to extend the idea of sustainable forms from the micro world to macroscopic scale and have a basis for a scientific foundation of concepts of natural ethics.

Conscious mental events are naturally correlated with events in human brains as they are conceived by quantum theory. Those events in the brain are typical Heisenberg events and their appearance is not governed by the known lows of physics. Decisions for them to occur or not to occur are the matter of the quantum choice. In words of Henry Stapp, such ontology offers a possibility for a meaningful Universe and a meaningful role of humans within it. Each actual thing (as an element of the physical reality in the micro world) is fundamentally the actualisation of an entire enduring complex macroscopic form. So, the micro world decides what happens in the macro world.

REFERENCES

Bassi, A., and Ghirardi, G., 2003, Dynamical reduction models, *Physics Reports* **379** (5-6): 257-426.

Bell, J. S., 1987, *Speakablee and Unspeakable in Quantum Mechanics*, Cambridge University Press, Cambridge.

Bohr, N., 1934, *Atomic Theory and the Description of Nature*, Cambridge University Press, Cambridge.

Bohr, N., 1963, *Essays on Atomic Physics and Human Knowledge*, Wiley, New York.

Boscivich, R. J., 1763, *Theoria Philosophiae Naturalis*, Venice.

Heisenberg, W., 1958, *Physics and Philosophy*, Harper and Row, New York, Chapter 3.

James, W., 1890/1950, *The Principles of Phychology*, Dover, New York, reprinted text from the original edition in 1890, p. 146.

Penrose, R., 1994, *Shadows of the Mind – An Approach to the Missing Science of Consciousness*, Oxford University Press, Oxford.

Stapp, H., 1993, *Mind, Matter and Quantum Mechanics*, Springer-Verlag, Berlin, Heidelberg, New York, Chapters 2 and 3.

Von Neumann, J., 1955, *Mathematical Foundations of Quantum Theory*, Princeton University Press, Princeton, Chapter 1.

Wigner, E., 1962, *The Scientific Speculates*, Basic Books, New York.

Wigner, E., 1967, Remarks on the mind-body questions, in: *Symmetries and Reflections*, Indiana University Press.

Zovko, N., 2001, Time in the mathematical theory of nature, *Synthesis Philosophica* **32**: 191-199.